REMARKS

Status of the Claims

Claims 1, 3, 5-11, 13-15, 17 and 18 are currently pending in the application. Claims 1, 3-11, 13-15 and 17 stand rejected. Claims 5-10, 13 and 14 have been withdrawn in response to a prior restriction requirement. Claims 1 has been amended as set forth herein. Claims 2, 4, 12 and 16 are cancelled. All cancellations and amendments are made without prejudice or disclaimer. No new matter has been added by way of the present amendments. Specifically, the amendment to claim 1 is supported by the specification at page 20, lines 10-15 (claim 1, part (e)), page 9, lines 8-12 (claim 1, part (e)), abstract (claim 1, part (f)), page 19, lines 14-20 and Figure 5 (claim 1, part (f) and part (g)), page 18, line 16 to page 19, line 1 (claim 1, part (h), and new claim 18) and page 9, lines 21-23 (claim 1, part (i)). Reconsideration is respectfully requested.

Interview

Applicants and Applicants' representatives thank the Examiner for extending the courtesy of an Interview on March 21, 2006. The substance of the interview is substantially as described in the Interview Summary of record. Briefly, amendments to claim 1, presented herein, were discussed, directed at further differentiating the present invention over the prior art and further clarifying that which Applicants consider to be their invention, addressing the written description issues. It is believed that the amendments and information discussed at the interview fully address all of the remaining issues and a Notice of Allowance is earnestly solicited.

Information Disclosure Statement

Applicants respectfully request the Examiner consider the references submitted in the Information Disclosure Statements of December 19, 2002 and April 10, 2003. If the references have already been considered, Applicants respectfully request the Examiner return to Applicants signed PTO-1449 forms indicating the references have been considered.

Rejections Under 35 U.S.C. § 112, First Paragraph

Claims 1, 3, 4, 11, 15 and 17 stand rejected under 35 U.S.C. § 112, first paragraph, for failing to comply with the written description requirement. (*See*, Office Action of November 18, 2005, at page 2, hereinafter referred to as "Office Action"). Claim 4 has been cancelled herein without prejudice or disclaimer, thus obviating the rejection as to claim 4. Applicants traverse the rejection as to the remaining claims as set forth herein.

The Examiner states the claims "broadly state an aspartic enzyme that produces plasma protein fragments having an inhibitory activity of metastasis and growth of cancer," and that "the claimed product is not isolated, nor described by any particular structure." (*Id.* at page 3). The Examiner further states that a "simple statement of function for the claimed protein does not ful[fill] the written description requirement." (*Id.*). Although Applicants do not agree that claim 1 lacks written description support in the specification, to expedite prosecution, claim 1 has been amended to recite the following additional features: (e) cleaves plasminogen at 73L-74F and/or 451L-452P to produce fragments comprising Kringles 1 to 4 of plasminogen; (f) is an aspartic protease; (g) has an activity that is inhibited by an aspartic protease inhibitor; (h) is isolated from mammalian cells by binding to an affinity chromatography column comprising an aspartic

MSW/TJS/py

protease inhibitor as a ligand; and (i) is Plasminogen Angiostatin Converting Enzyme at pH 4 (PACE4). Support for this amendment may be found in the as-filed specification at page 20, lines 10-15 (claim 1, part (e)), page 9, lines 8-12 (claim 1, part (e)), abstract (claim 1, part (f)), page 19, lines 14-20 and Figure 5 (claim 1, part (f) and part (g)), page 18, line 16 to page 19, line 1 (claim 1, part (h)) and page 9, lines 21-23 (claim 1, part (i)).

Claim 1 recites more than just a broad statement of an aspartic enzyme. The enzyme is defined functionally in claim 1 by precisely defining the enzyme's substrate and product. Furthermore, the name of the enzyme and its protease activity is recited in claim 1, at parts (e), (f) and (i). Therefore, claim 1 is clearly supported by the as-filed specification and one of ordinary skill in the art, having considered the as-filed specification, would know that Applicants were in possession of the presently claimed invention.

Reconsideration and withdrawal of the written description rejection of claims 1, 3, 4, 11, 15 and 17 are respectfully requested.

Rejections Under 35 U.S.C. § 102(b)

Claims 1, 3, 4, 11, 15 and 17 stand rejected under 35 U.S.C. § 102(b) as being anticipated by Gately et al., *Cancer Res.*, 56:4887-4890, 1996 (hereinafter referred to as "Gately et al."). (*See, Id.*, at page 4). Claim 4 has been cancelled herein without prejudice or disclaimer, thus obviating the rejection as to claim 4. Applicants traverse the rejection as to the remaining claims as set forth herein.

The Examiner states that "Gately's factor derived from PC-3 would in fact contain the claimed enzyme," and that "while Gately seemingly did not perform the proteinase inhibitor

MSW/TJS/py

assay at the same pH as Applicants this does not teach away from the simple fact that [the] enzyme of Gately contained in the PC-3 supernatent is the enzyme of Applicants." (*Id.*, at pages 4-5, emphasis added). However, in fact, it is biochemically impossible for the enzyme of Gately et al. to be the enzyme of the presently claimed invention.

The Enzyme of the Present Invention

As discussed at the interview, the present enzyme is an aspartic protease. (*See*, specification, as-filed, at abstract, page 19, lines 14-20 and Figure 5). Aspartic proteases are acidic proteases, meaning they belong to the family of proteases that operate at an acidic pH. (*See*, Rao et al., *Microbiol. Molecul. Biol. Rev.*, 62(3):597-635, 1998, at page 602-604, copy attached hereto as Exhibit A). The acidic proteases function most efficiently at an acidic pH. (*Id.* at page 602). Aspartic proteases, as exemplified by, for instance, pepsin, a very well-studied aspartic protease, catalyze peptide hydrolysis via active site aspartic acid residues. (*Id.* at page 603-604 and especially Figure 3(A)).

Aspartic proteases are inhibited with an exceptionally high specificity by pepstatin. (*See*, Dinu et al., *Roum. Biotechnol. Lett.*, 7(3):753-758, 2002, at abstract and pages 756 and 757, copy of reference attached hereto as Exhibit B, *see also*, Barrett et al., *Biochem. J.*, 127:439-441, 1972, copy attached hereto as Exhibit C). Pepstatin is a very well-characterized molecule, having been studied for over 30 years. (*Id.*). It is well known that pepstatin only inhibits, in a very specific manner, aspartic proteases. (*Id.*).

One of ordinary skill in the art would immediately conclude that the enzyme of the presently claimed invention can only be an aspartic protease based on the results of the data

MSW/TJS/py

disclosed in Figure 5 of the as-filed specification. Specifically, Figure 5 depicts the results of

inhibition studies wherein the enzyme of the present invention was incubated with various types

of protease enzyme inhibitors. Starting at the very top of the Figure 5, the lane having no

enzyme, the negative control, exhibits no activity (no solid bar extending to the right). However,

the enzyme alone, in the presence of no inhibitors, the positive control, exhibits activity as

indicated by the solid bar extending to the right. When the enzyme of the present invention is

incubated with various known and well-characterize protease inhibitors, only pepstatin

specifically and totally inhibited the activity of the enzyme of the presently claimed invention.

(See, Figure 5, lane marked "Pepstatin A", of the as-filed specification). Therefore, the enzyme

of the presently claimed invention can only be an aspartic protease and cannot be any other kind

of protease, such as, for instance, a serine protease.

The Enzyme Disclosed by Gately et al.

Serine proteases, in contrast, have entirely different enzymological mechanisms. (See,

Rao et al., at page 603, and Stryer, "Biochemistry", 3rd Edition, W.H. Freeman & Co., N.Y.,

1988, pages 220-226, copy of pages attached hereto as Exhibit D). Serine proteases, like the

protease disclosed in Gately et al., and like chymotrypsin, a very well studied member of the

serine protease family, have an enzymatic mechanism that utilizes a triad of active site amino

acid residues consisting of serine, histidine and aspartic acid. (See, Stryer, at page 224, Figure 9-

41). The mechanism of serine proteases is believed to include a tetrahedral transition state, as

depicted in Figure 9-44 of Stryer, at page 225. This is in marked contrast to the mechanism of

aspartic proteases which require the presence of two aspartic acid residues and one molecule of water in the active site to achieve peptide bond hydrolysis.

The enzyme disclosed by Gately et al. is only a serine protease and cannot also be an aspartic protease. As shown in Table 1 of Gately et al., pepstatin has no effect on the activity of the Gately et al. enzyme. However, several serine protease-specific inhibitors completely inhibit the activity of the enzyme of Gately et al. (*See*, Gately et al., at Table 1 and paragraph bridging pages 4888-4889). Thus, one of ordinary skill in the art would know that the enzyme of Gately et al. can only be a serine protease, and would also know that the enzyme of Gately et al. cannot be an aspartic protease.

The mechanisms of serine proteases and aspartic protease have been studied for many years. Chemical compounds that specifically inhibit a serine protease, but not any other types of proteases, have also been known for many decades. Likewise, chemical compounds, such as pepstatin, have been known for decades and are known to only inhibit aspartic proteases. The reason these chemical compounds are so exceptionally specific to inhibiting a particular kind of protease lies in the unique enzymatic catalytic mechanism of each family of protease. Thus, because the enzymatic mechanism of serine proteases is so different from that of aspartic proteases, an inhibitor of serine proteases is often not able to inhibit aspartic proteases.

Therefore, it is enzymatically impossible for the enzyme of the presently claimed invention to be a serine protease or to be the protease disclosed by Gately et al. because, as described in Table 1 of Gately et al. and Figure 5 of the present specification, the enzyme of the presently claimed invention is specifically inhibited by pepstatin, whereas that of Gately et al. is

not. Thus, the enzyme disclosed by Gately et al. must be a different enzyme and the disclosure of Gately et al. cannot anticipate the presently claimed invention.

Furthermore, the enzyme of the present invention is not only obtainable from PC-3 culture supernatants (human prostate cancer cells). The enzyme of the present invention is also obtainable from colon cancer cells (COLON) and lung cancer cells (LL/2). (*See*, as-filed specification, at Example 6 starting at page 40, Figure 10, page 33, lines 19-21 and at page 54). Thus, in contrast to the enzyme of Gately et al., which is only isolatable from PC-3 cells, the presently claimed enzyme is isolatable from multiple, different cell lines.

Dependent claims 3, 4, 11, 15 and 17 are not anticipated as, *inter alia*, depending from a non-anticipated base claim, claim 1.

Reconsideration and withdrawal of the anticipation rejection of claims 1, 3, 4, 11, 15 and 17 are respectfully requested.

CONCLUSION

If the Examiner has any questions or comments, please contact Thomas J. Siepmann,

Ph.D., Registration No 57,374 at the offices of Birch, Stewart, Kolasch & Birch, LLP.

If necessary, the Commissioner is hereby authorized in this, concurrent, and future

replies, to charge payment or credit any overpayment to our Deposit Account No. 02-2448 for

any additional fees required under 37 C.F.R. § 1.16 or under § 1.17; particularly, extension of

time fees.

Dated: April 18, 2006

Respectfully submitted,

By _____ #28977

Marc S. Weiner
Registration No.: 32,181
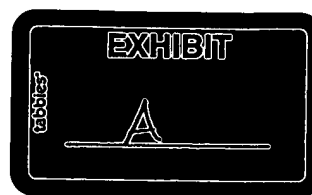BIRCH, STEWART, KOLASCH & BIRCH, LLP
8110 Gatehouse Road
Suite 100 East
P.O. Box 747
Falls Church, Virginia 22040-0747
(703) 205-8000
Attorney for Applicant

Attachments: Exhibits A-D

MSW/TJS/py

Molecular and Biotechnological Aspects of Microbial Proteases†

MALA B. RAO, APARNA M. TANKSALE, MOHINI S. GHATGE, AND VASANTI V. DESHPANDE*

Division of Biochemical Sciences, National Chemical Laboratory, Pune 411008, India

* Corresponding author. Mailing address: Division of Biochemical
Sciences, National Chemical Laboratory, Pune-411008, India. Phone:
091-212-338234. Fax: 091-212-338234.

INTRODUCTION

Proteases are the single class of enzymes which occupy a pivotal position with respect to their applications in both physiological and commercial fields. Proteolytic enzymes catalyze the cleavage of peptide bonds in other proteins. Proteases are degradative enzymes which catalyze the total hydrolysis of proteins. Advances in analytical techniques have demonstrated that proteases conduct highly specific and selective modifications of proteins such as activation of zymogenic forms of enzymes by limited proteolysis, blood clotting and lysis of fibrin clots, and processing and transport of secretory proteins across the membranes. The current estimated value of the worldwide sales of industrial enzymes is $1 billion (72). Of the industrial enzymes, 75% are hydrolytic. Proteases represent one of the three largest groups of industrial enzymes and account for about 60% of the total worldwide sale of enzymes (Fig. 1). Proteases execute a large variety of functions, extending from the cellular level to the organ and organism level, to produce cascade systems such as hemostasis and inflammation. They
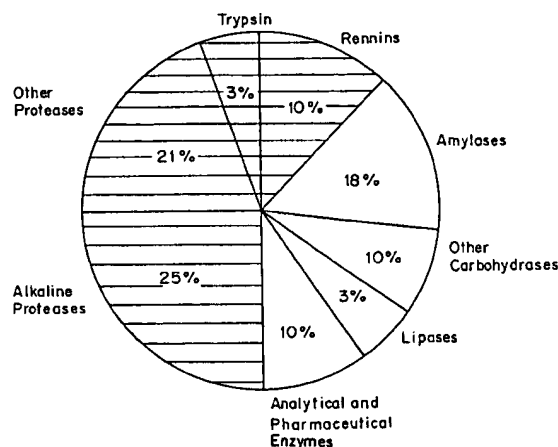
FIG. 1. Distribution of enzyme sales. The contribution of different enzymes to the total sale of enzymes is indicated. The shaded portion indicates the total sale of proteases.

are responsible for the complex processes involved in the normal physiology of the cell as well as in abnormal pathophysiological conditions. Their involvement in the life cycle of disease-causing organisms has led them to become a potential target for developing therapeutic agents against fatal diseases such as cancer and AIDS. Proteases have a long history of application in the food and detergent industries. Their application in the leather industry for dehairing and bating of hides to substitute currently used toxic chemicals is a relatively new development and has conferred added biotechnological importance (235). The vast diversity of proteases, in contrast to the specificity of their action, has attracted worldwide attention in attempts to exploit their physiological and biotechnological applications (64, 225). The major producers of proteases worldwide are listed in Table 1.

SCOPE OF THE REVIEW

Since proteases are enzymes of metabolic as well as commercial importance, there is a vast literature on their biochemical and biotechnological aspects (64, 128, 192, 235, 309). However, the earlier reviews did not deal with the molecular biology of proteases, which offers new possibilities and potentials for their biotechnological applications. This review aims at analyzing the updated information on biochemical and genetic aspects of proteases, with special reference to some of the advances made in these areas. We also attempt to address some of the deficiencies in the earlier reviews and to identify problems, along with possible solutions, for the successful applications of proteases for the benefit of mankind. The genetic engineering approaches are also discussed, from the perspective of making better use of proteases. The reference to plant and animal proteases has been made to complete the overview.

TABLE 1. Major protease producers

Company	Country	Market share (%)
Novo Industries	Denmark	40
Gist-Brocades	Netherlands	20
Genencor International	United States	10
Miles Laboratories	United States	10
Others		20

TABLE 2. Specificity of proteases

Enzyme	Peptide bond cleaved[a]
Trypsin	-Lys (or Arg) $\downarrow$----
Chymotrypsin, subtilisin	-Trp (or Tyr, Phe, Leu) $\downarrow$----
Staphylococcus V8 protease	-Asp (or Glu) $\downarrow$----
Papain	-Phe (or Val, Leu)-Xaa $\downarrow$----
Thermolysin	--- $\downarrow$ Leu (or Phe) ------
Pepsin	-Phe (or Tyr, Leu) $\downarrow$ Trp (or Phe, Tyr)

[a] The arrow indicates the site of action of the protease. Xaa, any amino acid residue.

However, the major emphasis of the review is on the microbial proteases.

SOURCES OF PROTEASES

Since proteases are physiologically necessary for living organisms, they are ubiquitous, being found in a wide diversity of sources such as plants, animals, and microorganisms.

Plant Proteases

The use of plants as a source of proteases is governed by several factors such as the availability of land for cultivation and the suitability of climatic conditions for growth. Moreover, production of proteases from plants is a time-consuming process. Papain, bromelain, keratinases, and ficin represent some of the well-known proteases of plant origin.

Papain. Papain is a traditional plant protease and has a long history of use (250). It is extracted from the latex of *Carica papaya* fruits, which are grown in subtropical areas of west and central Africa and India. The crude preparation of the enzyme has a broader specificity due to the presence of several proteinase and peptidase isozymes. The performance of the enzyme depends on the plant source, the climatic conditions for growth, and the methods used for its extraction and purification. The enzyme is active between pH 5 and 9 and is stable up to 80 or 90°C in the presence of substrates. It is extensively used in industry for the preparation of highly soluble and flavored protein hydrolysates.

Bromelain. Bromelain is prepared from the stem and juice of pineapples. The major supplier of the enzyme is Great Food Biochem., Bangkok, Thailand. The enzyme is characterized as a cysteine protease and is active from pH 5 to 9. Its inactivation temperature is 70°C, which is lower than that of papain.

Keratinases. Some of the botanical groups of plants produce proteases which degrade hair. Digestion of hair and wool is important for the production of essential amino acids such as lysine and for the prevention of clogging of wastewater systems.

Animal Proteases

The most familiar proteases of animal origin are pancreatic trypsin, chymotrypsin, pepsin, and rennins (23, 97). These are prepared in pure form in bulk quantities. However, their production depends on the availability of livestock for slaughter, which in turn is governed by political and agricultural policies.

Trypsin. Trypsin (M_r 23,300) is the main intestinal digestive enzyme responsible for the hydrolysis of food proteins. It is a serine protease and hydrolyzes peptide bonds in which the carboxyl groups are contributed by the lysine and arginine residues (Table 2). Based on the ability of protease inhibitors to inhibit the enzyme from the insect gut, this enzyme has received attention as a target for biocontrol of insect pests.

Trypsin has limited applications in the food industry, since the protein hydrolysates generated by its action have a highly bitter taste. Trypsin is used in the preparation of bacterial media and in some specialized medical applications.

Chymotrypsin. Chymotrypsin (M_r 23,800) is found in animal pancreatic extract. Pure chymotrypsin is an expensive enzyme and is used only for diagnostic and analytical applications. It is specific for the hydrolysis of peptide bonds in which the carboxyl groups are provided by one of the three aromatic amino acids, i.e., phenylalanine, tyrosine, or tryptophan. It is used extensively in the deallergenizing of milk protein hydrolysates. It is stored in the pancreas in the form of a precursor, chymotrypsinogen, and is activated by trypsin in a multistep process.

Pepsin. Pepsin (M_r 34,500) is an acidic protease that is found in the stomachs of almost all vertebrates. The active enzyme is released from its zymogen, i.e., pepsinogen, by autocatalysis in the presence of hydrochloric acid. Pepsin is an aspartyl protease and resembles human immunodeficiency virus type 1 (HIV-1) protease, responsible for the maturation of HIV-1. It exhibits optimal activity between pH 1 and 2, while the optimal pH of the stomach is 2 to 4. Pepsin is inactivated above pH 6.0. The enzyme catalyzes the hydrolysis of peptide bonds between two hydrophobic amino acids.

Rennin. Rennet is a pepsin-like protease (rennin, chymosin; EC 3.4.23.4) that is produced as an inactive precursor, prorennin, in the stomachs of all nursing mammals. It is converted to active rennin (M_r 30,700) by the action of pepsin or by its autocatalysis. It is used extensively in the dairy industry to produce a stable curd with good flavor. The specialized nature of the enzyme is due to its specificity in cleaving a single peptide bond in κ-casein to generate insoluble para-κ-casein and C-terminal glycopeptide.

Microbial Proteases

The inability of the plant and animal proteases to meet current world demands has led to an increased interest in microbial proteases. Microorganisms represent an excellent source of enzymes owing to their broad biochemical diversity and their susceptibility to genetic manipulation. Microbial proteases account for approximately 40% of the total worldwide enzyme sales (72). Proteases from microbial sources are preferred to the enzymes from plant and animal sources since they possess almost all the characteristics desired for their biotechnological applications.

Bacteria. Most commercial proteases, mainly neutral and alkaline, are produced by organisms belonging to the genus *Bacillus*. Bacterial neutral proteases are active in a narrow pH range (pH 5 to 8) and have relatively low thermotolerance. Due to their intermediate rate of reaction, neutral proteases generate less bitterness in hydrolyzed food proteins than do the animal proteinases and hence are valuable for use in the food industry. Neutrase, a neutral protease, is insensitive to the natural plant proteinase inhibitors and is therefore useful in the brewing industry. The bacterial neutral proteases are characterized by their high affinity for hydrophobic amino acid pairs. Their low thermotolerance is advantageous for controlling their reactivity during the production of food hydrolysates with a low degree of hydrolysis. Some of the neutral proteases belong to the metalloprotease type and require divalent metal ions for their activity, while others are serine proteinases, which are not affected by chelating agents.

Bacterial alkaline proteases are characterized by their high activity at alkaline pH, e.g., pH 10, and their broad substrate specificity. Their optimal temperature is around 60°C. These properties of bacterial alkaline proteases make them suitable for use in the detergent industry.

Fungi. Fungi elaborate a wider variety of enzymes than do bacteria. For example, *Aspergillus oryzae* produces acid, neutral, and alkaline proteases. The fungal proteases are active over a wide pH range (pH 4 to 11) and exhibit broad substrate specificity. However, they have a lower reaction rate and worse heat tolerance than do the bacterial enzymes. Fungal enzymes can be conveniently produced in a solid-state fermentation process. Fungal acid proteases have an optimal pH between 4 and 4.5 and are stable between pH 2.5 and 6.0. They are particularly useful in the cheesemaking industry due to their narrow pH and temperature specificities. Fungal neutral proteases are metalloproteases that are active at pH 7.0 and are inhibited by chelating agents. In view of the accompanying peptidase activity and their specific function in hydrolyzing hydrophobic amino acid bonds, fungal neutral proteases supplement the action of plant, animal, and bacterial proteases in reducing the bitterness of food protein hydrolysates. Fungal alkaline proteases are also used in food protein modification.

Viruses. Viral proteases have gained importance due to their functional involvement in the processing of proteins of viruses that cause certain fatal diseases such as AIDS and cancer. Serine, aspartic, and cysteine peptidases are found in various viruses (236). All of the virus-encoded peptidases are endopeptidases; there are no metallopeptidases. Retroviral aspartyl proteases that are required for viral assembly and replication are homodimers and are expressed as a part of the polyprotein precursor. The mature protease is released by autolysis of the precursor. An extensive literature is available on the expression, purification, and enzymatic analysis of retroviral aspartic protease and its mutants (147). Extensive research has focused on the three-dimensional structure of viral proteases and their interaction with synthetic inhibitors with a view to designing potent inhibitors that can combat the relentlessly spreading and devastating epidemic of AIDS.

Thus, although proteases are widespread in nature, microbes serve as a preferred source of these enzymes because of their rapid growth, the limited space required for their cultivation, and the ease with which they can be genetically manipulated to generate new enzymes with altered properties that are desirable for their various applications.

CLASSIFICATION OF PROTEASES

According to the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology, proteases are classified in subgroup 4 of group 3 (hydrolases) (114a). However, proteases do not comply easily with the general system of enzyme nomenclature due to their huge diversity of action and structure. Currently, proteases are classified on the basis of three major criteria: (i) type of reaction catalyzed, (ii) chemical nature of the catalytic site, and (iii) evolutionary relationship with reference to structure (12).

Proteases are grossly subdivided into two major groups, i.e., exopeptidases and endopeptidases, depending on their site of action. Exopeptidases cleave the peptide bond proximal to the amino or carboxy termini of the substrate, whereas endopeptidases cleave peptide bonds distant from the termini of the substrate. Based on the functional group present at the active site, proteases are further classified into four prominent groups, i.e., serine proteases, aspartic proteases, cysteine proteases, and metalloproteases (85). There are a few miscellaneous proteases which do not precisely fit into the standard classification, e.g., ATP-dependent proteases which require ATP for activity (183). Based on their amino acid sequences,

TABLE 3. Classification of proteases

Protease	Mode of action[a]	EC no.
Exopeptidases		
Aminopeptidases	●⌐O-O-O-O---	3.4.11
Dipeptidyl peptidase	●-●⌐O-O-O---	3.4.14
Tripeptidyl peptidase	●-●-●⌐O-O---	3.4.14
Carboxypeptidase	---O-O-O-O-O⌐●	3.4.16–3.4.18
Serine type protease		3.4.16
Metalloprotease		3.4.17
Cysteine type protease		3.4.18
Peptidyl dipeptidase	---O-O-O-O⌐●-●	3.4.15
Dipeptidases	●⌐●	3.4.13
Omega peptidases	⋆-●⌐O-O---	3.4.19
	---O-O-O⌐●-⋆	3.4.19
Endopeptidases	----O-O-O⌐O-O-O---	3.4.21–3.4.34
Serine protease		3.4.21
Cysteine protease		3.4.22
Aspartic protease		3.4.23
Metalloprotease		3.4.24
Endopeptidases of unknown catalytic mechanism		3.4.99

[a] Open circles represent the amino acid residues in the polypeptide chain. Solid circles indicate the terminal amino acids, and stars signify the blocked termini. Arrows show the sites of action of the enzyme.

proteases are classified into different families (5) and further subdivided into "clans" to accommodate sets of peptidases that have diverged from a common ancestor (236). Each family of peptidases has been assigned a code letter denoting the type of catalysis, i.e., S, C, A, M, or U for serine, cysteine, aspartic, metallo-, or unknown type, respectively.

Exopeptidases

The exopeptidases act only near the ends of polypeptide chains. Based on their site of action at the N or C terminus, they are classified as amino- and carboxypeptidases, respectively.

Aminopeptidases. Aminopeptidases act at a free N terminus of the polypeptide chain and liberate a single amino acid residue, a dipeptide, or a tripeptide (Table 3). They are known to remove the N-terminal Met that may be found in heterologously expressed proteins but not in many naturally occurring mature proteins. Aminopeptidases occur in a wide variety of microbial species including bacteria and fungi (310). In general, aminopeptidases are intracellular enzymes, but there has been a single report on an extracellular aminopeptidase produced by *A. oryzae* (150). The substrate specificities of the enzymes from bacteria and fungi are distinctly different in that the organisms can be differentiated on the basis of the profiles of the products of hydrolysis (31). Aminopeptidase I from *Escherichia coli* is a large protease (400,000 Da). It has a broad pH optimum of 7.5 to 10.5 and requires Mg^{2+} or Mn^{2+} for optimal activity (48). The *Bacillus licheniformis* aminopeptidase has a molecular weight of 34,000. It contains 1 g-atom of Zn^{2+} per mol, and its activity is enhanced by Co^{2+} ions. On the other hand, aminopeptidase II from *B. stearothermophilus* is a dimer with a molecular weight of 80,000 to 100,000 (272) and is activated by Zn^{2+}, Mn^{2+}, or Co^{2+} ions.

Carboxypeptidases. The carboxypeptidases act at C terminals of the polypeptide chain and liberate a single amino acid or a dipeptide. Carboxypeptidases can be divided into three major groups, serine carboxypeptidases, metallocarboxypeptidases, and cysteine carboxypeptidases, based on the nature of the amino acid residues at the active site of the enzymes. The

serine carboxypeptidases isolated from *Penicillium* spp., *Saccharomyces* spp., and *Aspergillus* spp. are similar in their substrate specificities but differ slightly in other properties such as pH optimum, stability, molecular weight, and effect of inhibitors. Metallocarboxypeptidases from *Saccharomyces* spp. (61) and *Pseudomonas* spp. (174) require Zn^{2+} or Co^{2+} for their activity. The enzymes can also hydrolyze the peptides in which the peptidyl group is replaced by a pteroyl moiety or by acyl groups.

Endopeptidases

Endopeptidases are characterized by their preferential action at the peptide bonds in the inner regions of the polypeptide chain away from the N and C termini. The presence of the free amino or carboxyl group has a negative influence on enzyme activity. The endopeptidases are divided into four subgroups based on their catalytic mechanism, (i) serine proteases, (ii) aspartic proteases, (iii) cysteine proteases, and (iv) metalloproteases. To facilitate quick and unambiguous reference to a particular family of peptidases, Rawlings and Barrett have assigned a code letter denoting the catalytic type, i.e., S, C, A, M, or U (see above) followed by an arbitrarily assigned number (236).

Serine proteases. Serine proteases are characterized by the presence of a serine group in their active site. They are numerous and widespread among viruses, bacteria, and eukaryotes, suggesting that they are vital to the organisms. Serine proteases are found in the exopeptidase, endopeptidase, oligopeptidase, and omega peptidase groups. Based on their structural similarities, serine proteases have been grouped into 20 families, which have been further subdivided into about six clans with common ancestors (12). The primary structures of the members of four clans, chymotrypsin (SA), subtilisin (SB), carboxypeptidase C (SC), and *Escherichia* D-Ala-D-Ala peptidase A (SE) are totally unrelated, suggesting that there are at least four separate evolutionary origins for serine proteases. Clans SA, SB, and SC have a common reaction mechanism consisting of a common catalytic triad of the three amino acids, serine (nucleophile), aspartate (electrophile), and histidine (base). Although the geometric orientations of these residues are similar, the protein folds are quite different, forming a typical example of a convergent evolution. The catalytic mechanisms of clans SE and SF (repressor LexA) are distinctly different from those of clans SA, SB, and SE, since they lack the classical Ser-His-Asp triad. Another interesting feature of the serine proteases is the conservation of glycine residues in the vicinity of the catalytic serine residue to form the motif Gly-Xaa-Ser-Yaa-Gly (25).

Serine proteases are recognized by their irreversible inhibition by 3,4-dichloroisocoumarin (3,4-DCI), L-3-carboxytrans 2,3-epoxypropyl-leucylamido (4-guanidine) butane (E.64), diisopropylfluorophosphate (DFP), phenylmethylsulfonyl fluoride (PMSF) and tosyl-L-lysine chloromethyl ketone (TLCK). Some of the serine proteases are inhibited by thiol reagents such as p-chloromercuribenzoate (PCMB) due to the presence of a cysteine residue near the active site. Serine proteases are generally active at neutral and alkaline pH, with an optimum between pH 7 and 11. They have broad substrate specificities including esterolytic and amidase activity. Their molecular masses range between 18 and 35 kDa, for the serine protease from *Blakeslea trispora*, which has a molecular mass of 126 kDa (76). The isoelectric points of serine proteases are generally between pH 4 and 6. Serine alkaline proteases that are active at highly alkaline pH represent the largest subgroup of serine proteases.

(i) Serine alkaline proteases. Serine alkaline proteases are produced by several bacteria, molds, yeasts, and fungi. They are inhibited by DFP or a potato protease inhibitor but not by tosyl-L-phenylalanine chloromethyl ketone (TPCK) or TLCK. Their substrate specificity is similar to but less stringent than that of chymotrypsin. They hydrolyze a peptide bond which has tyrosine, phenylalanine, or leucine at the carboxyl side of the splitting bond. The optimal pH of alkaline proteases is around pH 10, and their isoelectric point is around pH 9. Their molecular masses are in the range of 15 to 30 kDa. Although alkaline serine proteases are produced by several bacteria such as *Arthrobacter*, *Streptomyces*, and *Flavobacterium* spp. (21), subtilisins produced by *Bacillus* spp. are the best known. Alkaline proteases are also produced by *S. cerevisiae* (189) and filamentous fungi such as *Conidiobolus* spp. (219) and *Aspergillus* and *Neurospora* spp. (165).

(ii) Subtilisins. Subtilisins of *Bacillus* origin represent the second largest family of serine proteases. Two different types of alkaline proteases, subtilisin Carlsberg and subtilisin Novo or bacterial protease Nagase (BPN'), have been identified. Subtilisin Carlsberg produced by *Bacillus licheniformis* was discovered in 1947 by Linderstrom, Lang, and Ottesen at the Carlsberg laboratory. Subtilisin Novo or BPN' is produced by *Bacillus amyloliquefaciens*. Subtilisin Carlsberg is widely used in detergents. Its annual production amounts to about 500 tons of pure enzyme protein. Subtilisin BPN' is less commercially important. Both subtilisins have a molecular mass of 27.5 kDa but differ from each other by 58 amino acids. They have similar properties such as an optimal temperature of 60°C and an optimal pH of 10. Both enzymes exhibit a broad substrate specificity and have an active-site triad made up of Ser221, His64 and Asp32. The Carlsberg enzyme has a broader substrate specificity and does not depend on Ca^{2+} for its stability. The active-site conformation of subtilisins is similar to that of trypsin and chymotrypsin despite the dissimilarity in their overall molecular arrangements. The serine alkaline protease from the fungus *Conidiobolus coronatus* was shown to possess a distinctly different structure from subtilisin Carlsberg in spite of their functional similarities (218).

Aspartic proteases. Aspartic acid proteases, commonly known as acidic proteases, are the endopeptidases that depend on aspartic acid residues for their catalytic activity. Acidic proteases have been grouped into three families, namely, pepsin (A1), retropepsin (A2), and enzymes from pararetroviruses (A3) (13), and have been placed in clan AA. The members of families A1 and A2 are known to be related to each other, while those of family A3 show some relatedness to A1 and A2. Most aspartic proteases show maximal activity at low pH (pH 3 to 4) and have isoelectric points in the range of pH 3 to 4.5. Their molecular masses are in the range of 30 to 45 kDa. The members of the pepsin family have a bilobal structure with the active-site cleft located between the lobes (259). The active-site aspartic acid residue is situated within the motif Asp-Xaa-Gly, in which Xaa can be Ser or Thr. The aspartic proteases are inhibited by pepstatin (63). They are also sensitive to diazoketone compounds such as diazoacetyl-DL-norleucine methyl ester (DAN) and 1,2-epoxy-3-(p-nitrophenoxy)propane (EPNP) in the presence of copper ions. Microbial acid proteases exhibit specificity against aromatic or bulky amino acid residues on both sides of the peptide bond, which is similar to pepsin, but their action is less stringent than that of pepsin. Microbial aspartic proteases can be broadly divided into two groups, (i) pepsin-like enzymes produced by *Aspergillus*, *Penicillium*, *Rhizopus*, and *Neurospora* and (ii) rennin-like enzymes produced by *Endothia* and *Mucor* spp.

Cysteine/thiol proteases. Cysteine proteases occur in both prokaryotes and eukaryotes. About 20 families of cysteine proteases have been recognized. The activity of all cysteine proteases depends on a catalytic dyad consisting of cysteine and histidine. The order of Cys and His (Cys-His or His-Cys) residues differs among the families (12). Generally, cysteine proteases are active only in the presence of reducing agents such as HCN or cysteine. Based on their side chain specificity, they are broadly divided into four groups: (i) papain-like, (ii) trypsin-like with preference for cleavage at the arginine residue, (iii) specific to glutamic acid, and (iv) others. Papain is the best-known cysteine protease. Cysteine proteases have neutral pH optima, although a few of them, e.g., lysosomal proteases, are maximally active at acidic pH. They are susceptible to sulfhydryl agents such as PCMB but are unaffected by DFP and metal-chelating agents. Clostripain, produced by the anaerobic bacterium *Clostridium histolyticum*, exhibits a stringent specificity for arginyl residues at the carboxyl side of the splitting bond and differs from papain in its obligate requirement for calcium. Streptopain, the cysteine protease produced by *Streptococcus* spp., shows a broader specificity, including oxidized insulin B chain and other synthetic substrates. Clostripain has an isoelectric point of pH 4.9 and a molecular mass of 50 kDa, whereas the isoelectric point and molecular mass of streptopain are pH 8.4 and 32 kDa, respectively.

Metalloproteases. Metalloproteases are the most diverse of the catalytic types of proteases (13). They are characterized by the requirement for a divalent metal ion for their activity. They include enzymes from a variety of origins such as collagenases from higher organisms, hemorrhagic toxins from snake venoms, and thermolysin from bacteria (92, 210, 253, 311, 314). About 30 families of metalloproteases have been recognized, of which 17 contain only endopeptidases, 12 contain only exopeptidases, and 1 (M3) contains both endo- and exopeptidases. Families of metalloproteases have been grouped into different clans based on the nature of the amino acid that completes the metal-binding site; e.g., clan MA has the sequence HEXXH-E and clan MB corresponds to the motif HEXXH-H. In one of the groups, the metal atom binds at a motif other than the usual motif.

Based on the specificity of their action, metalloproteases can be divided into four groups, (i) neutral, (ii) alkaline, (iii) *Myxobacter* I, and (iv) *Myxobacter* II. The neutral proteases show specificity for hydrophobic amino acids, while the alkaline proteases possess a very broad specificity. *Myxobacter* protease I is specific for small amino acid residues on either side of the cleavage bond, whereas protease II is specific for lysine residue on the amino side of the peptide bond. All of them are inhibited by chelating agents such as EDTA but not by sulfhydryl agents or DFP.

Thermolysin, a neutral protease, is the most thoroughly characterized member of clan MA. Histidine residues from the HEXXH motif serve as Zn ligands, and Glu has a catalytic function (311). Thermolysin produced by *B. stearothermophilus* is a single peptide without disulfide bridges and has a molecular mass of 34 kDa. It contains an essential Zn atom embedded in a cleft formed between two folded lobes of the protein and four Ca atoms which impart thermostability to the protein. Thermolysin is a very stable protease, with a half-life of 1 h at 80°C.

Collagenase, another important metalloprotease, was first discovered in the broth of the anaerobic bacterium *Clostridium hystolyticum* as a component of toxic products. Later, it was found to be produced by the aerobic bacterium *Achromobacter iophagus* and other microorganisms including fungi. The action of collagenase is very specific; i.e., it acts only on collagen and
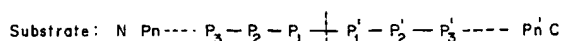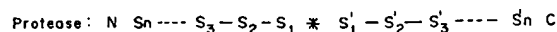
Protease: N   Sn ---- $S_3 - S_2 - S_1$ ∗ $S_1' - S_2' - S_3'$ ---- Sn   C

Substrate: N   Pn ---- $P_3 - P_2 - P_1$ ┼ $P_1' - P_2' - P_3'$ ---- Pn   C

FIG. 2. Active sites of proteases. The catalytic site of proteases is indicated by ∗ and the scissile bond is indicated by ┼; S1 through Sn and S1' through Sn' are the specificity subsites on the enzyme, while P1 through Pn and P1' through Pn' are the residues on the substrate accommodated by the subsites on the enzyme.

gelatin and not on any of the other usual protein substrates. Elastase produced by *Pseudomonas aeruginosa* is another important member of the neutral metalloprotease family.

The alkaline metalloproteases produced by *Pseudomonas aeruginosa* and *Serratia* spp. are active in the pH range from 7 to 9 and have molecular masses in the region of 48 to 60 kDa. *Myxobacter* protease I has a pH optimum of 9.0 and a molecular mass of 14 kDa and can lyse cell walls of *Arthrobacter crystellopoites*, whereas protease II cannot lyse the bacterial cells. Matrix metalloproteases play a prominent role in the degradation of the extracellular matrix during tissue morphogenesis, differentiation, and wound healing and may be useful in the treatment of diseases such as cancer and arthritis (26).

In summary, proteases are broadly classified as endo- or exoenzymes on the basis of their site of action on protein substrates. They are further categorized as serine proteases, aspartic proteases, cysteine proteases, or metalloproteases depending on their catalytic mechanism. They are also classified into different families and clans depending on their amino acid sequences and evolutionary relationships. Based on the pH of their optimal activity, they are also referred to as acidic, neutral, or alkaline proteases.

MECHANISM OF ACTION OF PROTEASES

The mechanism of action of proteases has been a subject of great interest to researchers. Purification of proteases to homogeneity is a prerequisite for studying their mechanism of action. Vast numbers of purification procedures for proteases, involving affinity chromatography, ion-exchange chromatography, and gel filtration techniques, have been well documented. Preparative polyacrylamide gel electrophoresis has been used for the purification of proteases from *Conidiobolus coronatus* (220). Purification of staphylocoagulase to homogeneity was carried out from culture filtrates of *Staphylococcus aureus* by affinity chromatography with a bovine prothrombin-Sepharose 4B column (109) and gel filtration (335). A number of peptide hydrolases have been isolated and purified from *E. coli* by DEAE-cellulose chromatography (217).

The catalytic site of proteases is flanked on one or both sides by specificity subsites, each able to accommodate the side chain of a single amino acid residue from the substrate. These sites are numbered from the catalytic site S1 through Sn toward the N terminus of the structure and S1' through Sn' toward the C terminus. The residues which they accommodate from the substrate are numbered P1 through Pn and P1' through Pn', respectively (Fig. 2).

Serine Proteases

Serine proteases usually follow a two-step reaction for hydrolysis in which a covalently linked enzyme-peptide intermediate is formed with the loss of the amino acid or peptide fragment (60). This acylation step is followed by a deacylation process which occurs by a nucleophilic attack on the intermediate by water, resulting in hydrolysis of the peptide. Serine

endopeptidases can be classified into three groups based mainly on their primary substrate preference: (i) trypsin-like, which cleave after positively charged residues; (ii) chymotrypsin-like, which cleave after large hydrophobic residues; and (iii) elastase-like, which cleave after small hydrophobic residues. The P1 residue exclusively dictates the site of peptide bond cleavage. The primary specificity is affected only by the P1 residues; the residues at other positions affect the rate of cleavage. The subsite interactions are localized to specific amino acids around the P1 residue to a unique set of sequences on the enzyme. Some of the serine peptidases from *Achromobacter* spp. are lysine-specific enzymes (179), whereas those from *Clostridium* spp. are arginine specific (clostripain) (71) and those from *Flavobacterium* spp. are post proline-specific (329). Endopeptidases that are specific to glutamic acid and aspartic acid residues have also been found in *B. licheniformis* and *S. aureus* (52).

The recent studies based on the three-dimensional structures of proteases and comparisons of amino acid sequences near the primary substrate-binding site in trypsin-like proteases of viral and bacterial origin suggest a putative general substrate binding scheme for proteases with specificity towards glutamic acid involving a histidine residue and a hydroxyl function. However, a few other serine proteases such as peptidase A from *E. coli* and the repressor LexA show distinctly different mechanism of action without the classic Ser-His-Asp triad (12). Some of the glycine residues are conserved in the vicinity of the catalytic serine residue, but their exact positions are variable (25).

The chymotrypsin-like enzymes are confined almost entirely to animals, the exceptions being trypsin-like enzymes from actinomycetes and *Saccharopolyspora* spp. and from the fungus *Fusarium oxysporum*.

A few of the serine proteases belonging to the subtilisin family show a catalytic triad composed of the same residues as in the chymotrypsin family; however, the residues occur in a different order (Asp-His-Ser). Some members of the subtilisin family from the yeasts *Tritirachium* and *Metarhizium* spp. require thiol for their activity. The thiol dependance is attributable to Cys173 near the active-site histidine (122).

The carboxypeptidases are unusual among the serine-dependent enzymes in that they are maximally active at acidic pH. These enzymes are known to possess a Glu residue preceding the catalytic Ser, which is believed to be responsible for their acidic pH optimum. Although the majority of the serine proteases contain the catalytic triad Ser-His-Asp, a few use the Ser-base catalytic dyad. The Glu-specific proteases display a pronounced preference for Glu-Xaa bonds over Asp-Xaa bonds (8).

Aspartic Proteases

Aspartic endopeptidases depend on the aspartic acid residues for their catalytic activity. A general base catalytic mechanism has been proposed for the hydrolysis of proteins by aspartic proteases such as penicillopepsin (121) and endothiapepsin (215). Crystallographic studies have shown that the enzymes of the pepsin family are bilobed molecules with the active-site cleft located between the lobes and each lobe contributing one of the pair of aspartic acid residues that is essential for the catalytic activity (20, 259). The lobes are homologous to one another, having arisen by gene duplication. The retropepsin molecule has only one lobe, which carries only one aspartic residue, and the activity requires the formation of a noncovalent homodimer (184). In most of the enzymes from the pepsin family, the catalytic Asp residues are contained in
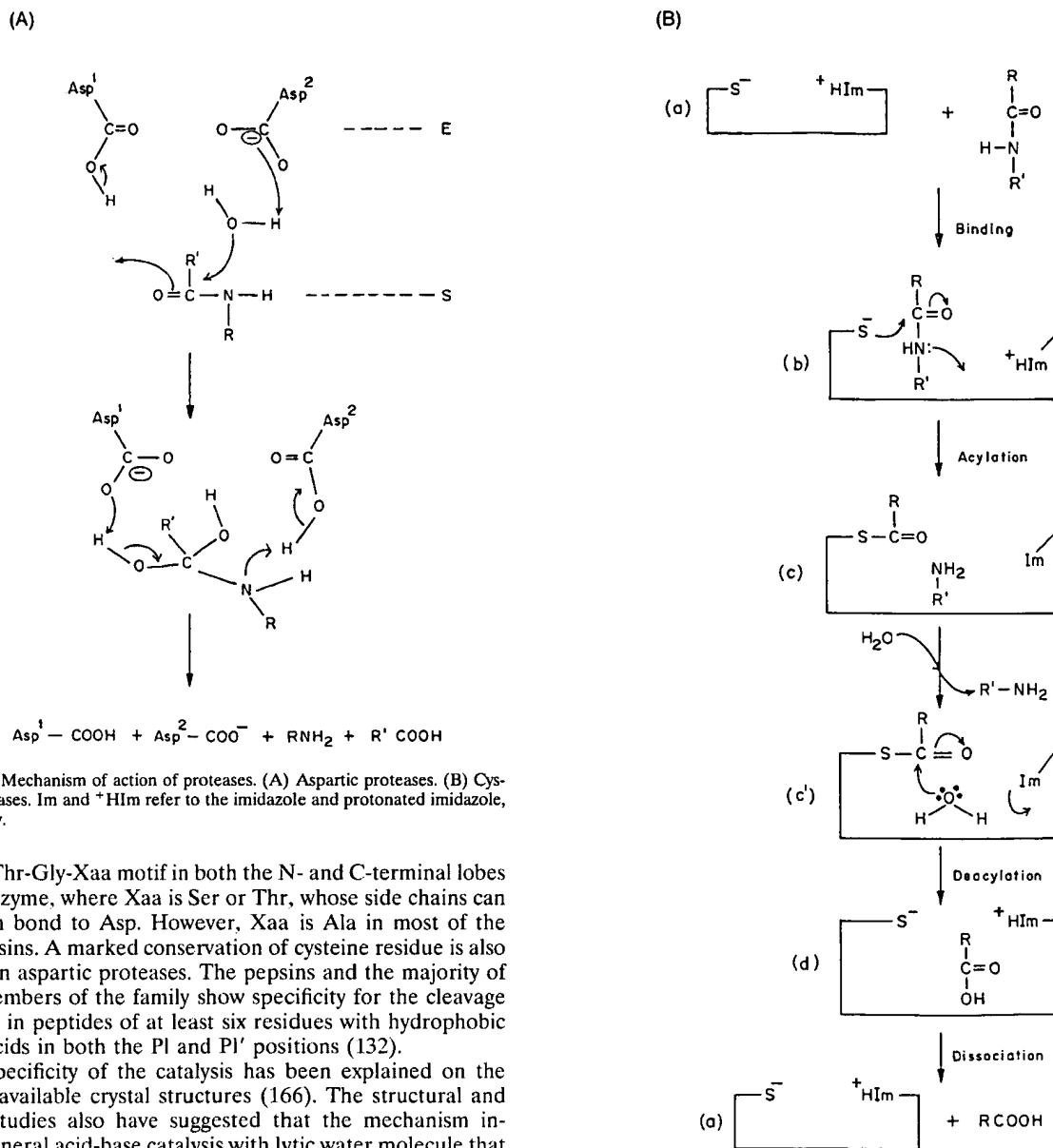
FIG. 3. Mechanism of action of proteases. (A) Aspartic proteases. (B) Cysteine proteases. Im and $^+$HIm refer to the imidazole and protonated imidazole, respectively.

an Asp-Thr-Gly-Xaa motif in both the N- and C-terminal lobes of the enzyme, where Xaa is Ser or Thr, whose side chains can hydrogen bond to Asp. However, Xaa is Ala in most of the retropepsins. A marked conservation of cysteine residue is also evident in aspartic proteases. The pepsins and the majority of other members of the family show specificity for the cleavage of bonds in peptides of at least six residues with hydrophobic amino acids in both the Pl and Pl' positions (132).

The specificity of the catalysis has been explained on the basis of available crystal structures (166). The structural and kinetic studies also have suggested that the mechanism involves general acid-base catalysis with lytic water molecule that directly participates in the reaction (Fig. 3A). This is supported by the crystal structures of various aspartic protease-inhibitor complexes and by the thiol inhibitors mimicking a tetrahedral intermediate formed after the attack by the lytic water molecule (120).

Metalloproteases

The mechanism of action of metalloproteases is slightly different from that of the above-described proteases. These enzymes depend on the presence of bound divalent cations and can be inactivated by dialysis or by the addition of chelating agents. For thermolysin, based on the X-ray studies of the complex with a hydroxamic acid inhibitor, it has been proposed that Glu143 assists the nucleophilic attack of a water molecule on the carbonyl carbon of the scissile peptide bond, which is polarized by the Zn^{2+} ion (98). Most of the metalloproteases are enzymes containing the His-Glu-Xaa-Xaa-His (HEXXH)

motif, which has been shown by X-ray crystallography to form a part of the site for binding of the metal, usually zinc.

Cysteine Proteases

Cysteine proteases catalyze the hydrolysis of carboxylic acid derivatives through a double-displacement pathway involving general acid-base formation and hydrolysis of an acyl-thiol intermediate. The mechanism of action of cysteine proteases is thus very similar to that of serine proteases.

A striking similarity is also observed in the reaction mechanism for several peptidases of different evolutionary origins. The plant peptidase papain can be considered the archetype of cysteine peptidases and constitutes a good model for this family of enzymes. They catalyze the hydrolysis of peptide, amide ester, thiol ester, and thiono ester bonds (226). The initial step in the catalytic process (Fig. 3B) involves the noncovalent

binding of the free enzyme (structure a) and the substrate to form the complex (structure b). This is followed by the acylation of the enzyme (structure c), with the formation and release of the first product, the amine R'-NH$_2$. In the next deacylation step, the acyl-enzyme reacts with a water molecule to release the second product, with the regeneration of free enzyme.

The enzyme papain consists of a single protein chain folded to form two domains containing a cleft for the substrate to bind. The crystal structure of papain confirmed the Cys25-His159 pairing (11). The presence of a conserved aspargine residue (Asn175) in the proximity of catalytic histidine (His159) creating a Cys-His-Asn triad in cysteine peptidases is considered analogous to the Ser-His-Asp arrangement found in serine proteases.

Studies of the mechanism of action of proteases have revealed that they exhibit different types of mechanism based on their active-site configuration. The serine proteases contain a Ser-His-Asp catalytic triad, and the hydrolysis of the peptide bond involves an acylation step followed by a deacylation step. Aspartic proteases are characterized by an Asp-Thr-Gly motif in their active site and by an acid-base catalysis as their mechanisms of action. The activity of metalloproteases depends on the binding of a divalent metal ion to a His-Glu-Xaa-Xaa-His motif. Cysteine proteases adopt a hydrolysis mechanism involving a general acid-base formation followed by hydrolysis of an acyl-thiol intermediate.

PHYSIOLOGICAL FUNCTIONS OF PROTEASES

Proteases execute a large variety of complex physiological functions. Their importance in conducting the essential metabolic and regulatory functions is evident from their occurrence in all forms of living organisms. Proteases play a critical role in many physiological and pathological processes such as protein catabolism, blood coagulation, cell growth and migration, tissue arrangement, morphogenesis in development, inflammation, tumor growth and metastasis, activation of zymogens, release of hormones and pharmacologically active peptides from precursor proteins, and transport of secretory proteins across membranes. In general, extracellular proteases catalyze the hydrolysis of large proteins to smaller molecules for subsequent absorption by the cell whereas intracellular proteases play a critical role in the regulation of metabolism. In contrast to the multitude of the roles contemplated for proteases, our knowledge about the mechanisms by which they perform these functions is very limited. Extensive research is being carried out to unravel the metabolic pathways in which proteases play an integral role; this research will continue to contribute significantly to our present state of information. Some of the major activities in which the proteases participate are described below.

Protein Turnover

All living cells maintain a particular rate of protein turnover by continuous, albeit balanced, degradation and synthesis of proteins. Catabolism of proteins provides a ready pool of amino acids as precursors of the synthesis of proteins. Intracellular proteases are known to participate in executing the proper protein turnover for the cell. In E. coli, ATP-dependent protease La, the lon gene product, is responsible for hydrolysis of abnormal proteins (38). The turnover of intracellular proteins in eukaryotes is also affected by a pathway involving ATP-dependent proteases (91). Evidence for the participation of proteolytic activity in controlling the protein turnover was

demonstrated by the lack of proper turnover in protease-deficient mutants.

Sporulation and Conidial Discharge

The formation of spores in bacteria (142), ascospores in yeasts (58), fruiting bodies in slime molds (205) and conidial discharge in fungi (221) all involve intensive protein turnover. The requirement of a protease for sporulation has been demonstrated by the use of protease inhibitors (41). Ascospore formation in yeast diploids was shown to be related to the increase in protease A activity (58). Extensive protein degradation accompanied the formation of a fruiting body and its differentiation to a stalk in slime molds. The alkaline serine protease of Conidiobolus coronatus was shown to be involved in forcible conidial discharge by isolation of a mutant with less conidial formation (221). Formation of the less active protease by autoproteolysis represents a novel means of physiological regulation of protease activity in C. coronatus (219).

Germination

The dormant spores lack the amino acids required for germination. Degradation of proteins in dormant spores by serine endoproteinases makes amino acids and nitrogen available for the biosynthesis of new proteins and nucleotides. These proteases are specific only for storage proteins and do not affect other spore proteins. Their activity is rapidly lost on germination of the spores (227). Microconidal germination and hyphal fusion also involve the participation of a specific alkaline serine protease (159). Extracellular acid proteases are believed to be involved in the breakage of cell wall polypeptide linkages during germination of Dictyostelium discoideum spores (118) and Polysphondylium pallidum microcysts (206).

Enzyme Modification

Activation of the zymogenic precursor forms of enzymes and proteins by specific proteases represents an important step in the physiological regulation of many rate-controlling processes such as generation of protein hormones, assembly of fibrils and viruses, blood coagulation, and fertilization of ova by sperm. Activation of zymogenic forms of chitin synthase by limited proteolysis has been observed in Candida albicans, Mucor rouxii, and Aspergillus nidulans. Kex-2 protease (kexin; EC 3.4.21.61), originally discovered in yeast, has emerged as a prototype of a family of eukaryotic precursor processing enzymes. It catalyzes the hydrolysis of prohormones and of integral membrane proteins of the secretory pathway by specific cleavage at the carboxyl side of pairs of basic residues such as Lys-Arg or Arg-Arg (12). Furin (EC 3.4.21.5) is a mammalian homolog of the Kex-2 protease that was discovered serendipitously and has been shown to catalyze the hydrolysis of a wide variety of precursor proteins at Arg-X-Lys or Arg-Arg sites within the constitutive secretory pathway (266). Pepsin, trypsin, and chymotrypsin occur as their inactive zymogenic forms, which are activated by the action of proteases.

Proteolytic inactivation of enzymes, leading to irreversible loss of in vivo catalytic activity, is also a physiologically significant event. Several enzymes are known to be inactivated in response to physiological or developmental changes or after a metabolic shift. Proteinases A and B from yeast inactivate several enzymes in a two-step process involving covalent modification of proteins as a marking mechanism for proteolysis.

Proteolytic modification of enzymes is known to result in a protein with altered physiological function; e.g., leucyl-L-RNA synthetase from E. coli is converted into an enzyme which

catalyzes leucine-dependent pyrophosphate exchange by removal of a small peptide from the native enzyme.

Nutrition

Proteases assist the hydrolysis of large polypeptides into smaller peptides and amino acids, thus facilitating their absorption by the cell. The extracellular enzymes play a major role in nutrition due to their depolymerizing activity. The microbial enzymes and the mammalian extracellular enzymes such as those secreted by pancreas are primarily involved in keeping the cells alive by providing them with the necessary amino acid pool as nutrition.

Regulation of Gene Expression

Modulation of gene expression mediated by protease has been demonstrated (241). Proteolysis of a repressor by an ATP-requiring protease resulted in a derepression of the gene. A change in the transcriptional specificity of the B subunit of *Bacillus thuringiensis* RNA polymerase was correlated with its proteolytic modification (154). Modification of ribosomal proteins by proteases has been suggested to be responsible for the regulation of translation (128).

Besides the general functions described so far, the proteases also mediate the degradation of a variety of regulatory proteins that control the heat shock response, the SOS response to DNA damage, the life cycle of bacteriophage (75), and programmed bacterial cell death (303). Recently, a new physiological function has been attributed to the ATP-dependent proteases conserved between bacteria and eukaryotes. It is believed that they act as chaperones and mediate not only proteolysis but also the insertion of proteins into membranes and the disassembly or oligomerization of protein complexes (275). In addition to the multitude of activities that are already assigned to proteases, many more new functions are likely to emerge in the near future.

APPLICATIONS OF PROTEASES

Proteases have a large variety of applications, mainly in the detergent and food industries. In view of the recent trend of developing environmentally friendly technologies, proteases are envisaged to have extensive applications in leather treatment and in several bioremediation processes. The worldwide requirement for enzymes for individual applications varies considerably. Proteases are used extensively in the pharmaceutical industry for preparation of medicines such as ointments for debridement of wounds, etc. Proteases that are used in the food and detergent industries are prepared in bulk quantities and used as crude preparations, whereas those that are used in medicine are produced in small amounts but require extensive purification before they can be used.

Detergents

Proteases are one of the standard ingredients of all kinds of detergents ranging from those used for household laundering to reagents used for cleaning contact lenses or dentures. The use of proteases in laundry detergents accounts for approximately 25% of the total worldwide sales of enzymes. The preparation of the first enzymatic detergent, "Burnus," dates back to 1913; it consisted of sodium carbonate and a crude pancreatic extract. The first detergent containing the bacterial enzyme was introduced in 1956 under the trade name BIO-40. In 1960, Novo Industry A/S introduced alcalase, produced by *Bacillus licheniformis*; its commercial name was BIOTEX. This

was followed by Maxatase, a detergent made by Gist-Brocades. The biggest market for detergents is in the laundry industry, amounting to a worldwide production of 13 billion tons per year. The ideal detergent protease should possess broad substrate specificity to facilitate the removal of a large variety of stains due to food, blood, and other body secretions. Activity and stability at high pH and temperature and compatibility with other chelating and oxidizing agents added to the detergents are among the major prerequisites for the use of proteases in detergents. The key parameter for the best performance of a protease in a detergent is its pI. It is known that a protease is most suitable for this application if its pI coincides with the pH of the detergent solution. Esperase and Savinase T (Novo Industry), produced by alkalophilic *Bacillus* spp., are two commercial preparations with very high isoelectric points (pI 11); hence, they can withstand higher pH ranges. Due to the present energy crisis and the awareness for energy conservation, it is desirable to use proteases that are active at lower temperatures. A combination of lipase, amylase, and cellulase is expected to enhance the performance of protease in laundry detergents.

All detergent proteases currently used in the market are serine proteases produced by *Bacillus* strains. Fungal alkaline proteases are advantageous due to the ease of downstream processing to prepare a microbe-free enzyme. An alkaline protease from *Conidiobolus coronatus* was found to be compatible with commercial detergents used in India (219) and retained 43% of its activity at 50°C for 50 min in the presence of Ca^{2+} (25 mM) and glycine (1 M) (16).

Leather Industry

Leather processing involves several steps such as soaking, dehairing, bating, and tanning. The major building blocks of skin and hair are proteinaceous. The conventional methods of leather processing involve hazardous chemicals such as sodium sulfide, which create problems of pollution and effluent disposal. The use of enzymes as alternatives to chemicals has proved successful in improving leather quality and in reducing environmental pollution. Proteases are used for selective hydrolysis of noncollagenous constituents of the skin and for removal of nonfibrillar proteins such as albumins and globulins. The purpose of soaking is to swell the hide. Traditionally, this step was performed with alkali. Currently, microbial alkaline proteases are used to ensure faster absorption of water and to reduce the time required for soaking. The use of nonionic and, to some extent, anionic surfactants is compatible with the use of enzymes. The conventional method of dehairing and dewooling consists of development of an extremely alkaline condition followed by treatment with sulfide to solubilize the proteins of the hair root. At present, alkaline proteases with hydrated lime and sodium chloride are used for dehairing, resulting in a significant reduction in the amount of wastewater generated. Earlier methods of bating were based on the use of animal feces as the source of proteases; these methods were unpleasant and unreliable and were replaced by methods involving pancreatic trypsin. Currently, trypsin is used in combination with other *Bacillus* and *Aspergillus* proteases for bating. The selection of the enzyme depends on its specificity for matrix proteins such as elastin and keratin, and the amount of enzyme needed depends on the type of leather (soft or hard) to be produced. Increased usage of enzymes for dehairing and bating not only prevents pollution problems but also is effective in saving energy. Novo Nordisk manufactures three different proteases, Aquaderm, NUE, and Pyrase, for use in soaking, dehairing, and bating, respectively.

Food Industry

The use of proteases in the food industry dates back to antiquity. They have been routinely used for various purposes such as cheesemaking, baking, preparation of soya hydrolysates, and meat tenderization.

Dairy industry. The major application of proteases in the dairy industry is in the manufacture of cheese. The milk-coagulating enzymes fall into three main categories, (i) animal rennets, (ii) microbial milk coagulants, and (iii) genetically engineered chymosin. Both animal and microbial milk-coagulating proteases belong to a class of acid aspartate proteases and have molecular weights between 30,000 to 40,000. Rennet extracted from the fourth stomach of unweaned calves contains the highest ratio of chymosin (EC 3.4.23.4) to pepsin activity. A world shortage of calf rennet due to the increased demand for cheese production has intensified the search for alternative microbial milk coagulants. The microbial enzymes exhibited two major drawbacks, i.e., (i) the presence of high levels of nonspecific and heat-stable proteases, which led to the development of bitterness in cheese after storage; and (ii) a poor yield. Extensive research in this area has resulted in the production of enzymes that are completely inactivated at normal pasteurization temperatures and contain very low levels of nonspecific proteases. In cheesemaking, the primary function of proteases is to hydrolyze the specific peptide bond (the Phe105-Met106 bond) to generate para-κ-casein and macropeptides. Chymosin is preferred due to its high specificity for casein, which is responsible for its excellent performance in cheesemaking. The proteases produced by GRAS (genetically regarded as safe)-cleared microbes such as *Mucor michei*, *Bacillus subtilis*, and *Endothia parasitica* are gradually replacing chymosin in cheesemaking. In 1988, chymosin produced through recombinant DNA technology was first introduced to cheesemakers for evaluation. Genencor International increased the production of chymosin in *Aspergillus niger* var. *awamori* to commercial levels. At present, their three recombinant chymosin products are available and are awaiting legislative approval for their use in cheesemaking (72).

Whey is a by-product of cheese manufacture. It contains lactose, proteins, minerals, and lactic acid. The insoluble heat-denatured whey protein is solubilized by treatment with immobilized trypsin.

Baking industry. Wheat flour is a major component of baking processes. It contains an insoluble protein called gluten, which determines the properties of the bakery doughs. Endo- and exoproteinases from *Aspergillus oryzae* have been used to modify wheat gluten by limited proteolysis. Enzymatic treatment of the dough facilitates its handling and machining and permits the production of a wider range of products. The addition of proteases reduces the mixing time and results in increased loaf volumes. Bacterial proteases are used to improve the extensibility and strength of the dough.

Manufacture of soy products. Soybeans serve as a rich source of food, due to their high content of good-quality protein. Proteases have been used from ancient times to prepare soy sauce and other soy products. The alkaline and neutral proteases of fungal origin play an important role in the processing of soy sauce. Proteolytic modification of soy proteins helps to improve their functional properties. Treatment of soy proteins with alcalase at pH 8 results in soluble hydrolysates with high solubility, good protein yield, and low bitterness. The hydrolysate is used in protein-fortified soft drinks and in the formulation of dietetic feeds.

Debittering of protein hydrolysates. Protein hydrolysates have several applications, e.g., as constituents of dietetic and health products, in infant formulae and clinical nutrition supplements, and as flavoring agents. The bitter taste of protein hydrolysates is a major barrier to their use in food and health care products. The intensity of the bitterness is proportional to the number of hydrophobic amino acids in the hydrolysate. The presence of a proline residue in the center of the peptide also contributes to the bitterness. The peptidases that can cleave hydrophobic amino acids and proline are valuable in debittering protein hydrolysates. Aminopeptidases from lactic acid bacteria are available under the trade name Debitrase. Carboxypeptidase A has a high specificity for hydrophobic amino acids and hence has a great potential for debittering. A careful combination of an endoprotease for the primary hydrolysis and an aminopeptidase for the secondary hydrolysis is required for the production of a functional hydrolysate with reduced bitterness.

Synthesis of aspartame. The use of aspartame as a noncalorific artificial sweetener has been approved by the Food and Drug Administration. Aspartame is a dipeptide composed of L-aspartic acid and the methyl ester of L-phenylalanine. The L configuration of the two amino acids is responsible for the sweet taste of aspartame. Maintenance of the stereospecificity is crucial, but it adds to the cost of production by chemical methods. Enzymatic synthesis of aspartame is therefore preferred. Although proteases are generally regarded as hydrolytic enzymes, they catalyze the reverse reaction under certain kinetically controlled conditions. An immobilized preparation of thermolysin from *Bacillus thermoprotyolyticus* is used for the enzymatic synthesis of aspartame. Toya Soda (Japan) and DSM (The Netherlands) are the major industrial producers of aspartame.

Pharmaceutical Industry

The wide diversity and specificity of proteases are used to great advantage in developing effective therapeutic agents. Oral administration of proteases from *Aspergillus oryzae* (Luizym and Nortase) has been used as a digestive aid to correct certain lytic enzyme deficiency syndromes. Clostridial collagenase or subtilisin is used in combination with broad-spectrum antibiotics in the treatment of burns and wounds. An asparginase isolated from *E. coli* is used to eliminate asparagine from the bloodstream in the various forms of lymphocytic leukemia. Alkaline protease from *Conidiobolus coronatus* was found to be able to replace trypsin in animal cell cultures (36).

Other Applications

Besides their industrial and medicinal applications, proteases play an important role in basic research. Their selective peptide bond cleavage is used in the elucidation of structure-function relationship, in the synthesis of peptides, and in the sequencing of proteins.

In essence, the wide specificity of the hydrolytic action of proteases finds an extensive application in the food, detergent, leather, and pharmaceutical industries, as well as in the structural elucidation of proteins, whereas their synthetic capacities are used for the synthesis of proteins.

GENETIC ENGINEERING OF MICROBIAL PROTEASES

Gene cloning is a rapidly progressing technology that has been instrumental in improving our understanding of the structure-function relationship of genetic systems. It provides an excellent method for the manipulation and control of genes. More than 50% of the industrially important enzymes are now produced from genetically engineered microorganisms (96).

TABLE 4. Cloning, sequencing, and/or expression of protease genes or cDNAs from microbial sources

Source of protease gene	Reference(s)	Source of protease gene	Reference(s)
Bacteria		*L. enzymogenes*	57
Bacilli		*Achromobacter lyticus* M497-1	208
B. subtilis 168		*A. lyticus*	169
apr	270	*Erwinia* sp.	2, 307
npr	90, 295, 297, 323	*Rhodocyclus gilatinosa* APR 3-2	116
epr	27, 263	*Bacteroids nodosus*	194
bpr	265	*Xanthomonas campestris* pv. campestris	168
mpr	264	*Treponema denticola* ATCC 33520	176
Isp-1	138	*Staphylococcus aureus* V8	29
B. subtilis (Natto) 16	319	*Thermus aquaticus* YT-1	148
B. subtilis N 515-N (*nprX*)	157	*Thermomonospora fusca* YX	152
Alkalophilic *Bacillus* strain	129	*Alteromonas* sp. strain O-7	298, 299
B. alkalophilus PB92	300	IgA family of proteases	
Bacillus sp. strain Y	293	*N. gonorrhoeae*	62, 224, 232
Alkalophilic *Bacillus* sp. NKS-21	318	*N. meningitidis*	169
Alkalophilic *Bacillus* sp. LG-12	251	*H. influenzae*	228
Bacillus sp. EA (Npr)	249	*Streptococcus sanguis* ATCC 10556	70
		S. pneumoniae	229, 308
Lactococci			
Streptococcus cremoris Wg2	139–141	**Fungi**	
Lactococcus lactis subsp. *cremoris* H2	317	Filamentous fungi	
Streptococcus lactis NCDO 763	137	Acidic proteases	
L. lactis subsp. cremoris SK11	50	*Mucor pusillus* rennin (MPR)	94, 296
L. lactis subsp. *lactis* VC 317	153	*Mucor miehei* aspartyl protease (MMAP)	51, 79
L. lactis subsp. *cremoris* Wg2	58	*R. niveus* aspartic protease (RNAP)	100, 101
Lactobacillus delbruckii subsp. bulgaricus	69	*A. awamori* aspergillopepsin A	15
		A. oryzae aspergillopepsin A	74
Streptomyces		*A. fumigatus* aspergillopepsin F	156, 237
S. griseus	89	*A. oryzae* M-9	284
S. griseus ATCC 10137	107	*A. satoi* ATCC 14332	257
S. cacaoi YM15	32	*A. niger* var. *macrosporus*	
S. fradiae ATCC 14544	135	Proctase A	114, 125, 283
S. lividans 66	17, 18, 163	Proctase B	113, 175
		Alkaline proteases (Alp)	
Serratia		*Aspergillus*	
Serratia sp. strain E-15	198	*A. oryzae* ATCC 20386	195, 207, 286, 288
S. marcescens SM6	24	*A. oryzae* Thailand industrial strain	33
S. marcescens	187	*A. soya*	211
S. marcescens ATCC 27117	134	*A. fumigatus*	123, 237
		A. flavus	233
Pseudomonas		*A. nidulans*	131
P. aeruginosa IFO 3455	7, 254	*Acremonium*	
P. aeruginosa PAO1	83	*A. chrysogenum* ATCC 11550	115
P. aeruginosa	82	*Fusarium*	136, 193
P. nalgiovense	68	Serine proteases	
		Tritirachium album Limber	
Aeromonas		Proteinase K	81
A. hydrophila SO 212	238, 239	Proteinase T	247
A. hydrophila D13	238	Metalloproteases	
		A. fumigatus MEP	124, 262
Vibrio		*A. flavus* MEP-20	234
V. anguillarum NB10	185	*A. fumigatus* MEP-20	234
V. parahaemolyticus	155		
V. vulnificus	34	**Yeasts**	
V. proteolyticus	43	Acidic proteases	
V. angionolyticus	45	*S. fibuligera* (PEP1)	95, 320
V. cholerae	86	*S. cerevisiae* (PEP4)	4, 315
		S. cerevisiae (BAR1)	177
E. coli		*S. cerevisiae* (YAP3)	53
Membrane proteases		*C. albicans* (SAP)	106, 170, 196
lspA, lep	42, 333	*C. tropicalis* (ACP)	294
sppA	108, 276	*Y. lipolytica* 148 (AXP)	331
ompT	80	Wild-type yeast	332
ATP-dependent proteases		Alkaline proteases	
La/Lon	3, 35, 334	*Y. lipolytica* (AEP) *XRP2*	44, 201
Clp/Ti	181	Serine proteases	
		Kluyveromyces lactis KEX-1	285
Miscellaneous		*S. cerevisiae* KEX-2	188, 212
Lysobacter enzymogenes 495	260		

TABLE 4—*Continued*

Source of protease gene	Reference(s)
Other proteases	
Yeast carboxypeptidase (CPY)	
S. cerevisiae PRC1	202
Vacoular protease B	
S. cerevisiae PRB1	190
Yeast proteasome *PRG1*	65
Viruses	
Animal viruses	
Herpesviruses	
HSV-1	47
HSV-2	271
MCMV	172
HHV-6	292
Adenoviruses	
Ad4	102
Ad12	103
Ad3	104
Ad40	306
Ad41	306
Retroviruses	
RSV	252
ASLV	144
ARV-2	248
M-MuLv	256
SRV-I	231
HTLV-2	255
BLV	245
M-PMV	105, 268
SIVmac	40
ARV	144
HTLV-I	246
HIV-1	46, 78, 171, 222
Picornaviruses	
Human rhinovirus type 14	162
Foot-and-mouth disease virus	6
Encephalomyocarditis virus	6
Poliovirus	6
Plant viruses	
Bean yellow mosaic virus	22
Zucchini yellow mosaic virus	
(Singapore isolate)	316

Several reports have been published in the past decade (Table 4) on the isolation and manipulation of microbial protease genes with the aim of (i) enzyme overproduction by the gene dosage effect, (ii) studying the primary structure of the protein and its role in the pathogenicity of the secreting microorganism, and (iii) protein engineering to locate the active-site residues and/or to alter the enzyme properties to suit its commercial applications. Protease genes from bacteria, fungi, and viruses have been cloned and sequenced (Table 4).

Bacteria

The objective of cloning bacterial protease genes has been mainly the overproduction of enzymes for various commercial applications in the food, detergent and pharmaceutical industries. The virulence of several bacteria is related to the secretion of several extracellular proteases. Gene cloning in these microbes was studied to understand the basis of their pathogenicity and to develop therapeutics against them. Proteases play an important role in cell physiology, and protease gene cloning, especially in *E. coli*, has been attempted to study the regulatory aspects of proteases.

Bacilli. (i) *B. subtilis* as a host for cloning of protease genes from *Bacillus* spp. The ability of *B. subtilis* to secrete various proteins into the culture medium and its lack of pathogenicity make it a potential host for the production of foreign polypeptides by recombinant DNA technology. Several *Bacillus* spp. secrete two major types of protease, a subtilisin or alkaline protease and a metalloprotease or neutral protease, which are of industrial importance. Studies of these extracellular proteases are significant not only from the point of view of overproduction but also for understanding their mechanism of secretion. Table 5 describes the cloning of genes for several neutral (*npr*) and alkaline (*apr*) proteases from various bacilli into *B. subtilis*.

(ii) *B. subtilis*. *B. subtilis* 168 secretes at least six extracellular proteases into the culture medium at the end of the exponential phase. The structural genes encoding the alkaline protease (*apr*) or subtilisin (270), neutral protease A and B (*nprA* and *nprB*) (90, 297, 323), minor extracellular protease (*epr*) (27, 263), bacillopeptidase F (*bpr*) (265), and metalloprotease (*mpr*) (264) have been cloned and characterized. These proteases are synthesized in the form of a "prepro" enzyme. To increase the expression of subtilisin and neutral proteases, Henner et al. replaced the natural promoters of *apr* and *npr* genes with the amylase promoter from *B. amyloliquefaciens* and the neutral protease promoter from *B. subtilis*, respectively (90). To understand the regulation of *npr* A gene expression, Toma et al. cloned the genes from *B. subtilis* 168 (normal producer) and Basc 1A341 (overproducer) (295). The two genes were found to be highly homologous except for a stretch of 66 bp close to the promoter region, which is absent in the Basc 1A341 gene. The *epr* gene shows partial homology to the *apr* gene and to the major intracellular serine protease (Isp-1) gene of *B. subtilis* (138). The *epr* gene was mapped at a locus different from the *apr* and *npr* loci on the *B. subtilis* chromosome and was shown not to be required for growth or sporulation, similar to *apr* or *npr* genes. Deletion of 240 amino acids (aa) from the C-terminal region of the *epr* gene product did not abolish the enzyme activity (27, 263). The deduced amino acid sequence of the mature *bpr* gene product is similar to those of other serine proteases of *B. subtilis*, i.e., subtilisin, Isp-1, and Epr. *B. subtilis* strains containing mutations in five extracellular protease genes (*apr*, *npr*, *epr*, *mpr*, and *bpr*) have been constructed (264) with the aim of expressing heterologous gene products in *B. subtilis*. The total amino acid sequence of *B. subtilis* Isp-1 deduced from the nucleotide sequence showed considerable homology (45%) to subtilisin. Highly conserved sequences are present around the essential amino acids, Ser, His, and Asp, indicating that the genes for both the intra- and extracellular serine proteases have a common ancestor.

In 1995, Yamagata et al. cloned and sequenced a 90-kDa serine protease gene (*hspK*) from *B. subtilis* (Natto) 16 (319). The large size of the enzyme may represent an ancient form of bacterial serine protease.

Analysis of DNA sequences of subtilisin BPN' from *B. amyloliquefaciens* (304, 313) and subtilisin Carlsberg from *B. licheniformis* (119) revealed that the two sequences are highly conserved in the coding region for the mature protein and must therefore have a common ancestral precursor. Yoshimoto et al. characterized the gene encoding subtilisin amylosacchariticus from *B. subtilis* subsp. *amylosacchariticus* (327, 328). The sequence was highly homologous to that of subtilisin E from *B. subtilis* 168 (269). The gene was expressed in *B. subtilis* ISW 1214 by using the vector pHY300PLK, with 20-

TABLE 5. Cloning of protease genes in *B. subtilis*

Source of proteases	Type of protease	Expression (fold)	Characterization of gene	Reference(s)
B. amyloliquefaciens F	Neutral	50	—[a]	99, 178
B. stearothermophilus F TELNE	Neutral	5	*nprS* sequenced; homologous to *nprM*	203
B. amyloliquefaciens IFO 14141	Neutral	15	Partially sequenced	133, 330
B. cereus	Neutral (metalloprotease)	√[b]	—	1
B. stearothermophilus CU21	Thermostable, neutral	√	*nprT* sequenced	67, 279
B. stearothermophilus 313-1	Thermostable, neutral	29	—	324
B. stearothermophilus HY-69	Thermostable, neutral	√	—	325
B. stearothermophilus MK232 and YG185-hyperproducing mutant of MK232	Highly thermostable, neutral		*nprM* sequenced; deduced amino acid sequence homologous to thermolysin (*B. thermoproteolyticus*) except for two substitutions, Asp37 to Asn37 and Glu119 to Gln119	145, 146
B. stearothermophilus	Thermostable, metalloprotease	√		258
B. brevis	Metalloprotease	—	√	9
B. licheniformis	Alkaline and neutral	√	—	209
B. amyloliquefaciens	Alkaline and neutral	—	*apr, npr* sequenced	304
B. licheniformis	Alkaline	√	—	88
B. pumilus IFO 12092	Alkaline	√	—	289
B. amyloliquefaciens	Subtilisin	√	√	313
B. natto	—	350	—	197
B. licheniformis ATCC 14580	C-terminal glutamic acid specific (BLase)	√	—	127

[a] —, no data is available.
[b] √, expression of the gene was observed.

fold-higher activity than that of the host and 4-fold-higher activity than that of *B. subtilis* subsp. *amylosacchariticus*.

(iii) **Alkalophilic *Bacillus* spp.** *Bacillus* proteases with an extremely alkaline pH optimum are generally used in detergent powders and are preferred over the subtilisins (optimal pH, 8.5 to 10.0). The information on these enzymes is helpful in designing new subtilisins. Kaneko et al. cloned and sequenced the *ale* gene, encoding alkaline elastase YaB, a new subtilisin from an alkalophilic *Bacillus* strain (129). The deduced amino acid sequence showed 55% homology to subtilisin BPN'. Almost all the positively charged residues have been predicted to be present on the surface of the alkaline elastase YaB molecule, facilitating its binding to elastin. The deduced amino acid sequence of the highly alkaline serine protease from another alkalophilic strain, *B. alcalophilus* PB92, showed considerable homology to YaB (300). The cloned gene was further used to increase the production level of the protease by gene amplification through chromosomal integration. Increased enzyme production and gene stabilization was observed when nontandem duplication occurred.

A gene encoding ISP-1 was characterized from alkalophilic *Bacillus* sp. strain NKS-21 (318). The nucleotide sequence was 50% homologous to genes encoding ISP-1 from *B. subtilis*, *B. polymyxa*, and the alkalophilic *Bacillus* sp. strain 221.

(iv) **Other bacilli.** A gene encoding the highly thermostable neutral proteinase (Npr) from *Bacillus* sp. strain EA1 was shown to be closely related to an *npr* gene from *B. caldolyticus* YP-T, except for a single-amino-acid change in the gene product (249). The enzyme from *Bacillus* sp. strain EA1 was more thermostable than the enzyme from *B. caldolyticus* YP-T; this can be attributed to the single-amino-acid change.

Lactococci. Lactococci (*Lactococcus lactis* subsp. *lactis* and *cremoris*, previously *Streptococcus lactis* and *Streptococcus cremoris*, respectively), the dairy starter cultures, have a complex proteolytic system which enables them to grow in milk by degrading casein into small peptides and free amino acids. This leads to the development of the texture and flavor of various dairy products. The importance of the cell envelope-located proteolytic system for dairy product quality has resulted in an increased fundamental research of the involved enzymes and their genes. On the basis of differences in caseinolytic specificity, the lactococcal proteases have been classified into two main groups: the PI-type protease, which degrades predominantly β-casein, and the PIII-type protease which degrades αS1-, β-, and κ-casein (305). Most of the genetic studies have focused on the PI-type protease genes. Lactococcal protease genes are located mostly on plasmids, which differ considerably in size and genetic organization in different strains (49). Curing experiments have suggested that plasmid pWV05 of *S. cremoris* Wg2 specifies proteolytic activity. The entire plasmid was subcloned in *E. coli* (140). A 4.3-MDa *Hin*dIII fragment of the plasmid, specifying the proteolytic activity, was cloned in *B. subtilis* and in a protease-deficient *S. lactis* strain. In *S. lactis*, the recombinant plasmid enabled the cells to grow normally in milk with rapid acid production. The *Hin*dIII fragment specifying the proteolytic activity of *S. cremoris* Wg2 was fully sequenced (141). The nucleotide sequence revealed two open reading frames (ORFs), ORF-1, a small ORF containing 295 codons, and ORF-2, a large ORF containing 1,772 codons. The protein specified by ORF-2 contained regions of extensive homology to subtilisins. The amino acids Asp32, His64, and Ser221, involved in the formation of the active site, were well

conserved. Deletion analysis of the proteinase gene of *S. cremoris* Wg2 showed that deletion of the C-terminal 343 aa did not influence the enzyme specificity of β-casein degradation (139). *L. lactis* subsp. *cremoris* H2 carries plasmid pDI21, containing the gene for the protease-positive phenotype (Prt⁺). The 6.5-kbp *Hin*dIII DNA fragment of pDI21 encoding the protease was cloned in *E. coli* as well as in *L. lactis* subsp. *lactis* 4125 (317). Protease that specifically degrades β-casein was expressed in both the transformed organisms. *S. lactis* NCDO 763 harbors plasmid pLP763, containing the gene for Prt⁺, which enables it to grow to a higher density in milk. The deduced amino acid sequence (1,902 aa) of the Prt⁺ phenotype was homologous to that of the serine protease from *S. cremoris* Wg2, suggesting that the genes encoding both products must have been derived from a common ancestral gene (137).

The PIII-type protease is found only in *L. lactis* subsp. *cremoris* AM1 and SK11. These strains are related, and they both contain the proteases encoded by the 78-kbp plasmid psk111. The *L. lactis* subsp. *cremoris* SK11 *prtP* gene was cloned and expressed in *E. coli* as well as in other subspecies of *L. lactis* (50). The location and orientation of the *prtP* gene on psk111 was determined by deletion analysis. A region at the C terminus of the *prtP* product, which is involved in cell envelope attachment, was identified. A deletion derivative of *prtP* specifying a C-terminally truncated protease was able to express and fully secrete the protease in the medium and showed the capacity to degrade αS1-, β-, and κ-casein. The N-terminal catalytic domain of the matrix enzyme shows significant sequence homology to the serine proteases of the subtilisin family (subtilases). Comparison with the known sequences of *prt* genes from *L. lactis* SK11, Wg2, and NCDO 763 indicated that the VC317 protease (153) is a natural hybrid of the SK11 and Wg2 proteases.

Stabilization of lactococcal protease genes (*prtP*, encoding the cell envelope-associated serine protease, and *prtM*, which activates the *prtP* gene product) is essential for the dairy industry. The plasmid-located *prtP* and *prtM* genes of *L. lactis* subsp. *cremoris* Wg2 were integrated (Campbell-like integration) into the *L. lactis* subsp. *lactis* MG1363 chromosome by using the insertion vector pKL9610 (158). Two transformants, MG610 and MG611, carrying different numbers (two and eight, respectively) of stable tandemly integrated plasmid copies, were obtained. Strain MG611 produced 11 times as much protease activity as did strain MG610 and about 1.5 times as much as did strain MG1363 (carrying five copies of the autonomously replicating plasmid).

A plasmid-free strain, *L. lactis* subsp. *cremoris* BC101, produces cell envelope-associated protease that is very similar or identical to the envelope protease encoded by the plasmid-linked *prtP* gene in other strains such as Wg2 and SK11. The *prtP* and *prtM* genes in this plasmid-free strain were identified on chromosomal DNA by pulsed-field gel electrophoresis (204). The chromosomal protease gene was shown to be organized in a fashion similar to that of the plasmid-linked protease gene. Recently, Gilbert et al. cloned and sequenced the *prtB* chromosomal gene from *Lactobacillus delbrueckii* subsp. *bulgaricus*, encoding a protease of 1,946 residues with a predicted molecular mass of 212 kDa (69). The deduced amino acid sequence showed significant homology to the N-terminal and catalytic domains of lactococcal PrtP cell surface proteases.

Streptomyces. *Streptomyces griseus*, an organism used for the commercial production of pronase, secretes two extracellular serine proteases: proteases A and B. The enzymes are 61% homologous on the basis of amino acid identity. The genes encoding protease A (*sprA*) and protease B (*sprB*) were isolated from the *S. griseus* genomic library, and their proteolytic

activity was demonstrated in *S. lividans* (89). The DNA sequences suggest that each protease is initially secreted as a precursor, which is then processed to remove an N-terminal propeptide from the mature protease. The strong homology between the coding regions of the two protease genes suggests that *sprA* and *sprB* must have originated by gene duplication. Protease B is one of the major extracellular proteases secreted by *S. griseus* ATCC 10137, and its gene was expressed in *S. lividans* by Hwang et al. (107). Their nucleotide sequencing of the gene further revealed that the deduced amino acid sequence was identical to that reported earlier by Henderson et al. (89). However, the nucleotide sequence of the 3'-flanking region was G rich and may be responsible for the reduced level of protease in *S. griseus* ATCC 10137 compared to the level in protease B-overproducing strains of *S. griseus*.

The *npr* gene for neutral metalloprotease from *S. cacaoi* YM15 was expressed in *S. lividans* (32). The deduced ORF encoded a 550-aa (60-kDa) protein, whereas the Npr secreted into the medium is 35 kDa, suggesting that it has undergone substantial processing since separating from the precursor.

S. fradiae ATCC 14544 secretes a novel, acidic-amino-acid-specific serine protease (SFase) into the culture medium. The deduced amino acid (135) sequence revealed a mature protein of 187 aa and shows 82% homology to the acidic-amino-acid-specific protease from *S. griseus* (277). Genes coding for a novel protease (163), a chymotrypsin-like serine protease (SAM-P20) (17), and SlpD and SlpE (homologs of the Tap [major tripeptidyl aminopeptidase] mycelium-associated proteases) (18) were cloned from *S. lividans* 66.

Serratia. The gram-negative bacteria belonging to the family *Enterobacteriaceae* are known to secrete large amounts of extracellular proteases into the surrounding medium. *Serratia* sp. strain E-15 produces a potent extracellular metalloprotease, which is widely used as an anti-inflammatory agent. The gene encoding the protease from *Serratia* sp. strain E-15 was expressed both in *E. coli* and in *S. marcescens* (198). Nucleotide sequence analysis revealed three zinc ligands (essential for proteolytic activity) and an active site, as predicted by comparing the deduced amino acid sequence with that of *B. thermoproteolyticus* thermolysin and *B. subtilis* neutral protease.

In another study, the extracellular serine protease (SSP) of *S. marcescens* was excreted through the outer membrane of *E. coli*. The nucleotide sequence of the cloned SSP gene, together with the determination of the N and C termini of the excreted enzymes, suggested that this protease is produced as a 112-kDa preproenzyme composed of an N-terminal signal sequence, the mature protease, and a large C-terminal domain (187).

Pseudomonas. *Pseudomonas aeruginosa* is an opportunistic pathogen and can cause fatal infections in compromised hosts. This virulence is related to the secretion of several extracellular proteins (167). *P. aeruginosa* secretes two proteases, an alkaline protease and an elastase. The alkaline protease genes (*apr*) from *P. aeruginosa* IFO 3455 and PAO1 were cloned in *E. coli* (7, 83, 254). The DNA fragment (8.8 kbp) coding for the alkaline protease from strain PAO1 was expressed in *E. coli* under the control of a *tac* promoter. Active enzyme was found to be synthesized and secreted into the medium in the absence of cell lysis.

The LasA protease (elastin degrading) of *P. aeruginosa* is also an important contributor to the pathogenesis of this bacterium. The enzyme shows a high level of staphylolytic activity. The *lasA* gene from strain FRD1 was overexpressed in *E. coli* (82). It encodes a precursor, prepro-LasA, of about 45 kDa. N-terminal sequence analysis allowed the identification of a 31-aa signal peptide. pro-LasA (42 kDa) does not undergo autoproteolytic processing and possesses little anti-staphylo-

coccal activity. The digestion of pro-LasA either by trypsin or by culture filtrate of the *P. aeruginosa lasA* deletion mutant yielded the active (20-kDa) staphylolytic protease.

Aeromonas. *Aeromonas hydrophila* and the related aeromonads are opportunistic pathogens of humans and fish. The pathogenicity of the microbe may involve several extracellular enzymes, and it has been suggested that the proteases excreted by *Aeromonas* spp. play an important role in invasiveness and in establishment of the infection. Two distinct types of extracellular proteases, a temperature-stable metalloprotease and a temperature-labile serine protease, are found in various strains of *A. hydrophila* and other aeromonads (160). Structural genes encoding extracellular proteases from two different *A. hydrophila* strains, SO2/2 and D13, were cloned in *E. coli* C600-1 by using pBR322 (238). A temperature-stable protease is secreted into the periplasm of *E. coli* and exhibits properties identical to those of the protease purified from *A. hydrophila* SO2/2 culture supernatant. A gene for the temperature-labile serine protease was also expressed from *A. hydrophila* SO2/2 into *E. coli* C600-1 and *S. lividans* 1326 (239).

Vibrio. To facilitate genetic analyses of the role of proteases in the pathogenesis of various *Vibrio* species, the genes encoding the Zn^{2+}-metalloprotease from *V. anguillarum* NB 10 (185), *V. parahaemolyticus* (155), and *V. vulnificus* (34) were cloned and sequenced. The conserved Zn^{2+}-binding domains were identified by measuring homology to other metalloproteases. The nucleotide sequence of the *nprV* gene encoding the extracellular neutral protease, vibriolysin (NprV), of *V. proteolyticus* revealed an ORF encoding 609 aa including a putative signal peptide sequence followed by a long prosequence of 172 aa (43). Comparative analysis of the mature NprV with the sequences of the neutral proteases from bacilli revealed extensive regions of conserved amino acid homology with respect to the active site and zinc- and calcium-binding residues. NprV was overproduced in *B. subtilis* by placing the DNA encoding the pro-NprV and the mature NprV downstream of the *Bacillus* promoter and signal sequences.

In one of the studies, the nucleotide sequence analysis of the structural gene, *hap*, for the extracellular haemagglutinin protease of *V. cholerae* revealed that the enzyme is produced as a large precursor, with the amino-terminal signal sequence following a propeptide (86). The deduced amino acid sequence of the mature enzyme showed 61.5% identity to the *P. aeruginosa* elastase.

E. coli. **(i) Membrane proteases.** In a bacterium, a protein that is to be exported across the cytoplasmic membrane is synthesized as a large precursor with a signal peptide at its amino terminus (19). The processing of this precursor involves two sequential events: (i) removal of the signal peptide from the precursor through an endo-type cleavage and (ii) digestion of the cleaved signal peptide. The membrane proteases involved are (i) signal peptidases (lipoprotein signal peptidase [Lsp] and leader peptidase [Lep]) and (ii) signal peptide peptidase (protease IV). The genes *lspA* (333), *lep* (42), and *sppA* (108, 276) for protease IV of *E. coli* have been characterized and mapped on *E. coli* chromosomal DNA. Protease IV was shown to be a tetramer of the *sppA* gene product.

(ii) ATP-dependent proteases. ATP-dependent proteolysis plays a major role in the turnover of both abnormal proteins and a variety of regulatory proteins in both prokaryotic and eukaryotic cells. Three families of ATP-dependent proteases are found in *E. coli*: La (or Lon), Clp (or Ti), and FtsH (or HflB) proteases. Lon and Clp are soluble proteins, whereas FtsH is a membrane-anchored protein.

In vitro studies on ATP-dependent proteolysis have shown that the major ATP-dependent activity in the extracts of *E. coli*

cells is the Lon protease (73). The *lon* gene of *E. coli* K-12 has been cloned (334), sequenced (3, 35), and shown to be dispensable by insertional mutagenesis of the gene (180). Extracts from Lon-deficient *E. coli* cells still catalyze ATP-dependent proteolysis mediated by a soluble two-component protease, Clp. Two dissimilar components of Clp are (i) the ClpA regulatory polypeptide, with two ATP-binding sites and an intrinsic ATPase activity, and (ii) the ClpP subunit, with a proteolytic active site. Clp is a serine protease, and its nucleotide sequence (181) showed little homology to the known classes of serine proteases representing a unique family of serine proteases (182).

The cleavage of proteins such as casein and albumin by Clp proteases requires both ClpP and the regulatory subunit ClpA and ATP. However, it has been observed that ClpP can independently catalyze endoproteolytic cleavage of short peptides at a lower rate than in the presence of ClpA and ATP. The gene encoding ClpP is, at 10 min on the *E. coli* map, nearer to the gene encoding the ATP-dependent Lon protease of *E. coli* and farther from the gene encoding ClpA. Primer extension experiments indicate that the transcription initiates immediately upstream of the coding region for ClpP, with a major transcription start at 120 bases in front of the start of translation. ClpP insertion mutants have been isolated, and strains devoid of ClpP are viable in the presence as well as the absence of Lon protease. Genetic evidence is available demonstrating that ClpA and ClpP act together in vivo (181). Processing of ClpP appears to involve an intermolecular autocatalytic cleavage reaction which is shown to be independent of ClpA (182). A speculative model for the chaperone-like function of ATP-dependent proteases has been postulated by Suzuki et al. (275). The dual function of the ATP-dependent protease is determined by the affinity of the protein for the subunit or domain. Based on this, the ATP-dependent protease may regulate the subunit stoichiometry of protein complexes.

Miscellaneous. Among the bacterial representatives of the trypsin family, α-lytic protease, an extracellular enzyme of the gram-negative soil bacterium *Lysobacter enzymogenes* 495, is of particular interest. Nucleotide sequence analysis and S1 mapping of the structural gene for the α-lytic protease from *L. enzymogenes* 495 indicated that the enzyme is synthesized as a prepro-protein (41 kDa) that is subsequently processed to its mature extracellular form (20 kDa) (260). The gene was further expressed in *E. coli* by fusing the promoter and signal sequence of the *E. coli phoA* gene to the proenzyme portion of the α-lytic protease gene (261). Following induction, an active enzyme was produced both intra- and extracellularly. Fusion of the mature protein domain alone resulted in the production of an inactive enzyme, indicating that the large N-terminal protein region is necessary for activity. Epstein and Wensink also cloned and sequenced the gene for α-lytic protease, a 19.8-kDa serine protease secreted by *L. enzymogenes* (57). The nucleotide sequence contains an ORF which codes for the 198-residue mature enzyme and a potential prepro-peptide, also of 198 residues.

Achromobacter protease I (API) is a mammalian-type, lysine-specific serine protease that specifically hydrolyzes the lysyl peptide bond. The nucleotide sequence analysis of API from *Achromobacter lyticus* M497-1 revealed that the gene codes for a single polypeptide chain of 653 aa (208). The 263-aa mature protein, which was identified by protein sequencing, was found to be flanked N-terminally by 205 aa including a signal peptide and C-terminally by 180 aa. *E. coli* carrying a recombinant plasmid containing the API gene overproduced and secreted the protein (API') into the periplasm. The N-terminal amino acid sequence of API' was the same as

that of mature API, whereas the enzyme retained the C-terminal extended polypeptide chain. The structural gene for β-lytic protease was cloned from *A. lyticus*, and the nucleotide sequence analysis of the gene revealed a mature enzyme of 179 aa, with additional 195 aa at the N-terminal end of the enzyme, which includes the signal peptide (161).

Characterization of a serine protease gene that cleaves specifically on the carbonyl side of acidic residues from *Staphylococcus aureus* V8 revealed a 68-residue N-terminal extension which includes a 19- to 29-residue signal peptide, the mature protein, and the C-terminal region with several repeated acidic amino acid-rich tripeptides (29). The C terminus may function as a competitive inhibitor of the prepro-protein form of the enzyme, perhaps to prevent activity prior to secretion.

Aqualysin I, an alkaline serine protease, is secreted into the culture medium by an extreme thermophile, *Thermus aquaticus* YT-1. Aqualysin I shows high DNA sequence homology to the subtilisin-type serine proteases, especially in the regions containing the active-site residues (Asp32, His64, and Ser221) of subtilisin BPN′ (148). The nucleotide sequence also revealed that the enzyme is produced as a large precursor, containing the N-terminal portion, the protease, and the C-terminal portion.

The gene (*tfgA*) for the major extracellular protease of *Thermomonospora fusca* YX was isolated, sequenced, and expressed in *Streptomyces lividans* (152). The ORF encoded 375 residues including a 31-residue potential signal sequence, an N-terminal 150-residue prosequence, and the 194-residue mature protease belonging to chymotrypsin family.

Alteromonas sp. strain O-7, a marine bacterium, excretes alkaline serine proteases or subtilases (AprI and AprII) into the growth medium. The results of the deduced amino acid sequence analysis of genes for both AprI and AprII indicated that both the enzymes are produced as large precursors consisting of four domains: the signal sequence, the N-terminal pro-region, the mature AprI or AprII, and the C-terminal extension (298, 299). The amino acid sequence of mature AprI shows high sequence homology to that of class I subtilase, while the sequence of AprII shows high sequence homology to that of class II subtilase. Repeated sequences were observed in the C-terminal pro-region, showing high homology to sequences from the C-terminal pro-region of other known gram-negative bacteria (*V. angiolyticum*, *Xanthomonas campestris*, and *V. proteolyticus*).

IgA family of proteases. Immunoglobulin A1 (IgA1) proteases form a very heterogenous group of extracellular endopeptidases produced by a number of bacterial pathogens that colonize human mucosal surfaces. The enzymes specifically cleave human IgA1, which participates in the immune system surveillance in the human mucosa. A number of reports (62, 224, 232) on the cloning of the *iga* gene, encoding the IgA1 protease from *Neisseria gonorrhoeae*, are available. Nucleotide sequence analysis revealed that the enzyme is produced as a large precursor with three functional domains, i.e., the N-terminal leader peptide, the protease, and the carboxy-terminal "helper" domain. An overall structural similarity to the *iga* gene from *N. meningitidis* was also demonstrated (169).

Comparison of the deduced amino acid sequence of the *iga* gene of *Haemophilus influenzae* serotype b with that of a similar protease from *N. gonorrhoeae* revealed several domains with a high degree of homology (228). An enzyme secretion mechanism analogous to that for *N. gonorrhoeae* IgA1 protease was proposed for *H. influenzae* IgA1 protease. Limited diversity has been found among the IgA1 protease genes of *H. influenzae*, serotype b strains (230), information that is useful from the point of view of vaccine preparation.

Cloning of streptococcal IgA1 genes from *Streptococcus sanguis* ATCC 10556 (70) and *S. pneumoniae* (229, 308) has been reported. Hybridization experiments with an *S. sanguis* IgA1 protease gene probe showed no detectable homology to chromosomal DNA of gram-negative bacteria secreting IgA1 proteases. The gene encoding IgA1 protease from *S. pneumoniae* was identified by using the *S. sanguis* protease probe. However, the *iga* gene was found to be highly heterogenous among streptococcal species.

From the foregoing, it can be seen that subtilisins (270) and neutral proteases (279, 323) of various *Bacillus* species, the α-lytic protease from *L. enzymogenes* (57, 260), and proteases A and B from *S. griseus* (89) have long polypeptide extensions at their N termini. The IgA protease of *N. gonorrhoeae* (224) and the protease of *S. marcescens* (322) have C-terminal extensions. *Achromobacter* protease I (208), aqualysin I from *T. aquaticus* (148), and AprI and AprII from *Alteromonas* sp. strain O-7 (298, 299) bear long peptide chains at both the N and C termini. The function of the pre-peptide portion (signal peptide) in these precursors is possibly to assist in the transport of the secretory protein across the cytoplasmic membrane. The exact role of the pro-peptide region is not known; possibly the long peptide serves to inhibit the mature protease to which it is connected (29, 57). It is also possible that the pro-peptide helps the protease to fold into its active form (111, 261).

Fungi

As in bacteria, cloning of the protease genes of fungi has been attempted from both the commercial and pathogenicity points of view.

Filamentous fungi. (i) Acidic proteases. *(a) Mucor.* Two closely related species of zygomycete fungus, *Mucor pusillus* and *Mucor miehei*, secrete aspartate proteases, also known as mucor rennins, into the medium. The enzymes possess high milk-clotting activity and low proteolytic activity, enabling them to be used as substitutes for calf chymosin in the cheese industry.

Sequencing of the cloned gene encoding *M. pusillus* rennin (MPR) revealed an ORF without introns, encoding possible pre-pro-sequences (66 aa) upstream of the mature MPR sequence (296). The deduced amino acid sequence showed a high degree of homology to that of *M. miehei* rennin (MMR). The gene encoding *M. miehei* aspartyl protease (MMAP) has also been cloned and sequenced (79). The deduced primary translation product showed an N-terminal extension which appears to comprise a signal peptide of 22 aa and a propeptide of 47 aa. Fungal aspartyl proteases are structurally related to each other and to the gastric aspartyl proteases chymosin and pepsin; therefore, they may be activated in a manner similar to their gastric counterparts. When the gene encoding the pre-pro-form of MPR was cloned in *S. cerevisiae* under the control of the yeast GAL7 promoter, an inactive zymogen of the enzyme with the 44-aa pro-sequence was identified in the medium during the initial stage of cultivation (94). In vitro conversion of the zymogen to mature MRP was shown to proceed autocatalytically under the acidic conditions.

(b) Rhizopus. *Rhizopus niveus*, belonging to the zygomycete class, also secretes aspartyl protease abundantly. The gene encoding *R. niveus* aspartic protease (RNAP) was cloned and sequenced (100). Comparison of the deduced amino acid sequence with the amino acid sequence of rhizopuspepsin of *R. chinensis* (282) revealed that the RNAP gene has an intron within its coding region. A prepro-sequence of 66 aa upstream of the mature enzyme was also revealed. High-level secretion of RNAP-I was achieved by subcloning the RNAP-I gene into

Saccharomyces cerevisiae (101). Yeast cells carrying the intact RNAP-I gene under the control of the glyceraldehyde-3-phosphate dehydrogenase gene promoter of *S. cerevisiae* were unable to synthesize RNAP-I. On removal of the intron of the RNAP-I gene, the cell secreted the enzyme with high efficiency.

(c) *Aspergillus*. The *pepA* gene encoding the aspartic protease, aspergillopepsin A, from *Aspergillus awamori* (15), the *pepA* gene from *A. oryzae* (74), and the cDNA coding for an elastinolytic aspartic protease, aspergillopepsin F, from *A. fumigatus* (156, 237) were cloned and sequenced. The nucleotide sequence data revealed that the ORFs encoding aspartic proteases in these aspergilli are composed of four exons. Prepropeptides of 69, 78, and 70 aa were found to precede 395-, 326-, and 323-aa mature proteins of *A. awamori*, *A. oryzae*, and *A. fumigatus*, respectively. The amino acid sequence of aspergillopepsin F shows 70, 66, and 67% homology to the sequences of those from *A. oryzae*, *A. awamori*, and *A. saitoi*, respectively. The primary structure of aspergillopepsin I from *A. satoi* ATCC 14332 (now designated *A. phoenicis*) was deduced from the nucleotide sequence of the gene (257). The cDNA of the gene was also cloned and expressed in yeast cells.

Two types of acid endopeptidases, acid proteases A and B (commercially named proctase A and B), are known to be secreted into the medium by *A. niger* var. *macrosporus*. Protease B is a typical aspartic protease, inhibited by pepstatin, whereas protease A is not inhibited by pepstatin. Sequencing of the protease A gene revealed an 846-bp structural gene without any introns encoding the precursor form of the enzyme (114, 125, 283). The precursor, of 282 residues, includes an N-terminal prepro-sequence of 59 residues, the L chain of 39 residues, an intervening sequence of 11 residues, and the H chain of 173 residues linked in that order. The deduced amino acid sequence (394 residues) of the prepro-form of protease B showed 98% homology to the sequences of aspergillopepsin I from *A. awamori* and *A. saitoi* and 68% homology to that of aspergillopepsin I from *A. oryzae* (113, 175). The cDNA was expressed in *E. coli*, and the purified pro-protease B showed protease activity under acidic conditions (pH 2 to 4).

(ii) **Alkaline proteases.** *(a)* *Aspergillus*. Alkaline protease (Alp) produced by *A. oryzae*, a filamentous ascomycete used in the manufacture of soy sauce, is considered to play an important role in producing the flavor of soy sauce by hydrolyzing the raw materials. Tatsumi et al. constructed the cDNA library of *A. oryzae* ATCC 20386 in pUC119 and isolated a cDNA (1,100 bp) encoding the mature region of Alp (286). The nucleotide sequence of the cDNA lacked most of the DNA sequences corresponding to the prepro-region. The entire cDNA coding for prepro-Alp was cloned and expressed in *S. cerevisiae* (288). The character of the Alp secreted from *S. cerevisiae* was shown to be identical to that of the native Alp. The predicted mature Alp consists of 282 aa and shows homology to other serine proteases of subtilisin families from bacteria as well as from fungi. Alp has a 121-aa prepro-region wherein the N-terminal 21 residues show the characteristics of a signal peptide. Alp expressed in *S. cerevisiae* was secreted with the N terminus processed correctly, analogous to the expression in *S. cerevisiae* of aspartic protease from *M. pusillus* (321). The prepro-Alp cDNA of *A. oryzae* was further cloned into an osmophilic yeast, *Zygosaccharomyces rouxii* (207). The recombinant *Z. rouxii* secreted a large amount of Alp (about 300 mg/liter) into the culture medium. The Alp gene is 1,374 nucleotides long and contains three introns, one in the pro-region and two in the mature protein region (195).

A gene encoding Alp from the *A. oryzae* Thailand industrial strain was isolated from the genomic library by using oligode-

oxyribonucleotide probes based on the *A. oryzae* ATCC 20386 cDNA sequence (33). By comparison with the published cDNA sequence (286), Alp from *A. oryzae* Thailand was found to be encoded by four exons. Transformation of the *alpA* gene in the high-level-Alp-producing *A. oryzae* strain U212, obtained by UV mutagenesis, resulted in the production of up to five times as much Alp as in the parental strain. *A. fumigatus* and *A. flavus*, the agents of invasive aspergillosis, secrete highly homologous serine proteases. The genomic as well as cDNA clones encoding elastinolytic Alp from both *A. fumigatus* (123, 237) and *A. flavus* (233) were sequenced. The *A. nidulans prtA* gene coding for Alp was isolated by using the gene encoding *A. oryzae* Alp (131). The nucleotide sequence of *prtA* was determined, and the deduced amino acid sequence showed a high degree of similarity to Alp from *A. fumigatus*, *A. flavus*, and *A. oryzae*. *prtA* transcription was shown to be dependent on the medium composition.

(b) *Acremonium*. *Acremonium chrysogenum* ATCC 11550 (*Cephalosporium acremonium*) produces a considerable amount of extracellular Alp. The cDNA and genomic DNA encoding Alp were isolated from the *A. chrysogenum* cDNA and genomic DNA libraries, respectively (115). The nucleotide sequence of the gene was determined. The deduced amino acid sequence showed 57% homology to that of *A. oryzae* Alp. Cloning of the entire cDNA encoding *A. chrysogenum* Alp into *S. cerevisiae* directed the secretion of enzymatically active Alp into the culture medium.

(c) *Fusarium*. The transfer of the *Fusarium* alkaline protease gene (136) into *A. chrysogenum* resulted in transformants producing large amount of Alp (193). Southern hybridization analysis, as well as PCR of genomic DNAs from these transformants, showed chromosomal integration of the full-length *alp* gene. The enzyme secreted by *A. chrysogenum* had properties identical to that of the native *Fusarium* Alp, indicating that the Alp promoter, signal sequence, and introns functioned correctly in *A. chrysogenum*.

(iii) **Serine proteases.** *(a)* *Tritirachium*. Proteinase K is a serine endoproteinase excreted by the fungus *Tritirachium album* Limber. The enzyme is able to hydrolyze native proteins rapidly and is active in the presence of detergents (urea, sodium dodecyl sulfate, etc.), making the proteinase K one of the most useful tools in molecular biology. The enzyme exhibits strong similarity to the bacterial subtilisins. The genomic DNA as well as the cDNA encoding proteinase K from *T. album* Limber have been cloned in *E. coli*, and the entire nucleotide sequence of the coding region, including the 5'- and 3'-flanking regions, has been determined (81). The nucleotide sequence analysis revealed that the primary secreted product is a zymogen containing a 15-aa signal sequence and a 90-aa pro-peptide. The pro-peptide is presumably removed in the later steps of the secretion process or upon secretion into the medium. The proteinase K gene was shown to be composed of two exons and one 63-bp intron located in the proregion. The pro-proteinase K gene was expressed in *E. coli* under the control of the *tac* promoter.

The coding sequence of proteinase T from *T. album* Limber (247) was shown to be interrupted by two introns. The deduced amino acid sequence showed 53% identity to that of proteinase K. The presence of four cysteines in the mature proteinase, probably in the form of two disulfide bonds, explains the thermal stability of proteinase T. The proteinase T cDNA was expressed in *E. coli*, and the authenticity of the product was confirmed by Western blotting and N-terminal analysis of the recombinant product.

(iv) **Metalloproteases.** *(a)* *Aspergillus*. Jaton-Ogay et al. (124) and Sirakova et al. (262) cloned and sequenced the gene

as well as the cDNA encoding the 42-kDa elastinolytic metalloproteinase (MEP) of *A. fumigatus*. Comparison of the nucleotide sequences revealed that the genomic and the cDNA sequences are analogous except for four introns interrupting the ORF. The enzyme was shown to be produced in a preproform, with a 384-aa mature protease region. In another study, no intron was found in the ORF of *A. flavus mep20* (encoding a 23-kDa MEP) whereas a 59-bp intron was present in the gene from *A. fumigatus* (a homolog of *mep20*) (234). The MEP20 proteins of *A. flavus* and *A. fumigatus* have 68% identity.

Yeasts. (i) Acidic proteases. The yeast *Saccharomycopsis fibuligera* produces an extracellular acid protease (PEP1). DNA coding for the secretable acid protease gene of *S. fibuligera* was isolated (95, 320). The enzyme produced by *Saccharomyces cerevisiae* cells that are transformed with a plasmid carrying the cloned gene showed enzymatic properties similar to those of the *S. fibuligera* protease.

Two different groups of workers (4, 315) from the United States worked simultaneously on the *PEP4* gene of *S. cerevisiae*, which encodes an aspartyl protease implicated in the posttranslational regulation of the yeast vacuolar hydrolases. The *PEP4* gene was isolated from a genomic library by complementation of the *PEP4-3* mutation. The nucleotide sequence was deduced, and the predicted amino acid sequence showed substantial homology to that of the aspartyl protease family.

The deduced primary translation product (587 aa) of *bar1*, the structural gene for the barrier activity of *S. cerevisiae*, has a putative signal peptide and nine potential asparagine-linked glycosylation sites (177). Marked sequence similarity to pepsinlike proteases was observed.

A gene for yeast aspartyl protease 3 (YAP3) allowing KEX-2-independent MFα pro-pheromone processing was isolated from *S. cerevisiae* (53). The nucleotide sequence of the YAP3-encoding gene was determined, and the deduced amino acid sequence was shown to exhibit extensive homology to a number of aspartyl proteases, including the PEP4 and BAR1 proteins of *S. cerevisiae*. A potential transmembrane domain similar to that found in the KEX-2 gene product was also located.

Candida albicans and *Candida tropicalis* are the medically more important opportunistic pathogens causing infections in immunocompromised patients. Their secretory proteolytic activity is considered to be a major virulence factor. The deduced amino acid sequence of the acid protease (ACP) from *C. tropicalis* shows similarity to the amino acid sequence of the pepsin family (294). The aspartyl proteinase gene (106, 170) and cDNA (196) from various *C. albicans* strains were cloned and sequenced. The genes for secreted aspartic proteases (the *SAP1*, *SAP2*, *SAP3*, and *SAP4* genes) in *C. albicans* constitute a multigene family. Three putative new members, *SAP5*, *SAP6*, and *SAP7*, were also isolated and sequenced. Evidence was also obtained for the existence of *SAP* multigene families in other *Candida* species such as *C. tropicalis*, *C. parapsilosis*, and *C. guilliermondii* (191).

The amino acid sequence of an acid extracellular protease (AXP) from *Yarrowia lipolytica* 148 deduced from the nucleotide sequence revealed a putative 17-aa pre-peptide, a 27-aa pro-peptide, and a 353-aa mature protein (37 kDa) (331). AXP showed homology to proteases of several fungal genera. The transcription of both AXP and the alkaline extracellular protease (AEP) genes in *Y. lipolytica* was shown to be regulated by the pH of the culture (331).

A gene encoding an extracellular protease was cloned from a wild-type yeast into brewer's yeast, *S. cerevisiae* (332). Such genetically engineered strains carrying the gene for an extracellular protease were shown to exhibit chill-proofing activity in beer. Proteins remaining in beer after its brewing from malt

tend to form hazes during chilling due to their poor solubility at lower temperatures. Acid proteases assist in reducing the haze formation by degrading the proteins in beer without affecting foam stability or organoleptic properties such as taste.

(ii) Alkaline protease. The *XRP2* gene for AEP from *Y. lipolytica* encodes a putative 22-aa pre-peptide followed by a 135-aa pro-peptide containing a possible N-linked glycosylation site and the two Lys-Arg peptidase-processing sites (44, 201). The mature protease (297 aa) contains two potential glycosylation sites.

(iii) Serine proteases. *(a) Kluyveromyces.* The *KEX-1* gene product is required for the production of a killer toxin by *Kluyveromyces lactis*. The deduced amino acid sequence (700 aa) encoded by *KEX-1* showed an internal domain with a striking homology to the sequences of the subtilisin-type proteinases (285).

(b) S. cerevisiae. The *KEX-2* gene, encoding a subtilisin-like endoprotease responsible for posttranslational processing of certain gene products, contains a 2,442-bp ORF encoding a polypeptide of 814 aa (188, 212). The deduced amino acid sequence revealed a region near the N terminus that has extensive homology to the subtilisin family of serine proteases. A putative membrane-spanning domain near the C terminus was also detected. The wild type and the C-terminal deletion derivatives showed similar substrate specificities, with the highest activity being against Arg-Arg dipeptides.

(iv) Other proteases. Yeast carboxypeptidase (CPY) is a glycosylated yeast vacuolar protease that is used commercially in peptide synthesis. CPY is encoded by the *PRC1* gene. To increase the production of CPY in *S. cerevisiae*, *PRC1* was placed under the control of the strongly regulated yeast GAL1 promoter on the multicopy plasmids and introduced into *vpl1* mutant strains (202). About a 200-fold increase in the level of secreted CPY (40 mg/liter) was obtained compared to the level in a *vpl1* mutant carrying a single copy of the wild-type *PRC-1* gene. Sodium dodecyl sulfate-polyacrylamide gel electrophoresis revealed two forms of secreted active CPY, probably due to the different levels of glycosylation. The structural gene *PRB1*, encoding the vacuolar protease B of *S. cerevisiae*, was cloned by complementation of the *prb1-1122* mutation (190).

PRG1, a yeast gene encoding the 32-kDa proteasome, which shows 55.6% sequence homology to 80% of the RING10 gene product (human proteasome), was identified (65). Genomic disruption of *PRG1* revealed that it is essential for yeast growth. The results strongly indicate that the antigen-processing system present in vertebrates has evolved from a basic cellular process present in all organisms.

Viruses

Gene cloning of viral proteases has been undertaken for the isolation and overexpression of the gene and for subsequent screening of inhibitory compounds that may be used in the development of chemotherapeutic agents. Viral protease is responsible for processing of polyprotein precursors into the structural proteins of the mature virion. Among viruses, reports on cloning of protease genes are limited mainly to animal viruses (Table 5).

Animal viruses. (i) Herpesviruses. Each member of the herpesvirus family encodes a unique serine protease in association with a capsid assembly protein, with the associated ORFs being designated U_L80 and U_L26 in human cytomegalovirus (HCMV) and in herpes simplex virus type-1 (HSV-1), respectively. The U_L26 gene encodes a protease responsible for the C-terminal cleavage of the nucleocapsid-associated proteins (ICP35c and ICP35d) to their posttranslationally modified

counterparts (ICP35e and ICP35f). The protease expressed in *E. coli* exhibited autoprocessing and specifically cleaved the ICP35 protein assembly (47). Similarly, genes encoding proteases from HSV-2, murine cytomegalovirus (MCMV), and human herpesvirus 6 (HHV-6) have been studied (172, 271, 292). Such studies assist in the investigation of the role of proteolytic processing in the virus.

(ii) Adenoviruses. Adenoviruses code for a serine-centered, neutral protease specific for selected Gly-Ala bonds in several virus-encoded precursor proteins that are required for virion maturation and infectivity. To determine the functional domains of this key enzyme, protease genes from various types of adenoviruses have also been cloned and sequenced (102–104, 306).

(iii) Retroviruses. The genomic organization of retroviruses is 5'-LTR-*gag-pro-pol-env*-LTR-3' (where LTR is a long terminal repeat). The *pro/prt* gene product is an aspartyl protease, which is responsible for processing the *gag* and *pol* polyprotein precursors into the structural proteins of the mature virion. Comparison of the genomic organization of certain retroviruses revealed that *prt* lies in the carboxyl terminus of *gag* in Rous sarcoma virus (RSV) (252) and avian sarcoma leukosis virus (ASLV) (144); in the amino terminus of *pol* in AIDS-associated retrovirus type 2 (ARV-2) (248); in the same reading frame as both *gag* and *pol* in Moloney murine leukemia virus (M-MuLV) (256); and as a separate reading frame in simian AIDS retrovirus type I (SRV-I) (231), human T-cell leukemia virus type 2 (HTLV-2) (255), bovine leukaemia virus (BLV) (245), and Mason-Pfizer monkey virus (MPMV) (268). Besides cloning and sequencing of the *prt* gene, there are a few reports on expression of the gene in *E. coli* (40, 105, 144). Significant inhibition of the expressed protease activity by pepstatin A confirmed that HTLV-1 protease is a member of the aspartyl protease group (246).

Human immunodeficiency virus (HIV), a causative agent of AIDS, is also a member of the family *Retroviridae*. The virus exhibits the same overall *gag-pol-env* genome organization as that of other retroviruses. The genome-size mRNA of HIV-1 is translated into two polyproteins: Pr55 (*gag* gene product) and Pr160 (*gag-pol* gene product). Cleavage of these polyproteins by the viral protease into smaller structural proteins and replication enzymes such as reverse transcriptase and integrase is necessary to produce infectious progeny virions from immature virus particles. The enzyme, a part of the polyprotein, has a highly conserved sequence, Asp-Thr-Gly, which is homologous to the active site of the aspartic proteases and is thought to belong to this enzyme family (216). The protease is essential for the retroviral life cycle, as indicated by the production of noninfectious, replication-deficient virions by Moloney murine leukemia virus variants mutated in the protease-encoding region (130). This suggests that HIV protease is a good target for chemotherapy and that specific inhibitors of this enzyme may have a significant function in the treatment of AIDS without interfering with the host cell physiology. To obtain sufficient quantities of the HIV protease for biochemical and structural analyses, several groups have described expression of the recombinant HIV-1 protease in *E. coli* (46, 78, 171). Pichuantes et al. have reported extracellular expression of HIV-1 aspartic protease in *S. cerevisiae* (222). The expressed enzyme was shown to exhibit a proteolytic activity, as has been shown to be associated with the purified HIV-1 virions (164). Debouck et al. expressed the HIV protease gene product in *E. coli* (46). The product was shown to autocatalyze its maturation from a larger precursor and to process an HIV Pr55 gag protein when coexpressed in *E. coli*. This allowed a structure-function analysis of the HIV protease and provided a simple assay for the

development of potential therapeutic agents directed against the critical viral enzyme.

(iv) Picornaviruses. Human rhinovirus is a member of the picornavirus (small RNA) family. Rhinovirus has commercial importance since it is the causative agent of about 15% of cases of the common cold. A cDNA encoding the viral protease from the 3C region of human rhinovirus type 14 was expressed in *E. coli* through the use of a periplasmic secretion vector (162). A comparison of the 3C protease regions of all the available picornavirus (foot-and-mouth disease virus, encephalomyocarditis virus, and poliovirus) sequences revealed two completely conserved residues, Cys147 and His161, which may be the reactive residues of the active sites of these cysteine proteases (6).

Plant viruses. Potyviruses are a cause of serious losses of several major crop plants. In plants infected with the potyviruses, inclusions consisting of viral proteins are found in the cell nucleus. One of them, the nuclear inclusion protease (NIa), is the major viral protein responsible for the proteolytic maturation of the polyprotein encoded by the virus. The elucidation of the structure of such virus-encoded proteins could eventually facilitate the design of novel polypeptides which bind to them and inhibit their functions. With this objective, cDNAs for NIa proteases were cloned and sequenced from bean yellow mosaic virus (22) and zucchini yellow mosaic virus (Singapore isolate) (316).

The potential contributions of genetic engineering to mankind are enormous and will benefit agriculture, animal husbandry, environmental protection, food production and processing, human health care, manufacture of biochemicals and biofuels, etc. In general, the application of genetic engineering to proteases will facilitate their use in industry and enable the development of therapeutic agents against the proteases that are important in the life cycle of organisms which cause serious diseases.

PROTEIN ENGINEERING

Many industrial applications of proteases require enzymes with properties that are nonphysiological. Protein engineering allows the introduction of predesigned changes into the gene for the synthesis of a protein with an altered function that is desired for the application. Recent advances in recombinant DNA technology and the ability to selectively exchange amino acids by site-directed mutagenesis (SDM) have been responsible for the rapid progress of protein engineering. Identification of the gene and knowledge of the three-dimensional structure of the protein in question are the two main prerequisites for protein engineering. The X-ray crystallographic structures of several proteases have been determined (39, 143, 223, 267). Proteases from bacteria, fungi, and viruses have been engineered to improve their properties to suit their particular applications.

Bacteria

Subtilisin has been chosen as a model system for protein engineering since a lot of basic information about this commercially important enzyme is available. Its pH dependance (290), catalytic activity (278, 281), stability to heat or denaturing agents (112, 199), and substrate specificity (10, 14, 30, 59, 243) have been altered through SDM. A slightly reduced rate of thermal inactivation was observed for a subtilisin BPN' variant containing two cysteine residues (Cys22, Cys87) (186, 214). Oxidation of Met222 adjacent to the Ser221 in the active site of subtilisin reduces the catalytic activity of subtilisin. The

effect of substitution of Met222 with different amino acids revealed that small side chains yield the highest activity. The mutant enzymes Ser222, Ala222, and Leu222 were active and stable to peroxide for 1 h. Probing of the specificity of the S_1 binding site of Met222 Cys/Ser mutants of subtilisin from *Bacillus lentus* with boronic acid inhibitors revealed similar binding trends for the mutant and the parent (269). The disulfide bonds introduced into subtilisin away from its catalytic center were shown to possess increased autolytic stability (312). Higher thermostability of subtilisin E as a result of introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease has been reported (280). Strausberg et al. have created the environment for stabilization of subtilisin by deleting the calcium-binding loop from the protein (273). Analysis of the structure and stability of the prototype with the loop deleted followed by SDM resulted in a mutant with native proteolytic activity and 1,000-fold-greater stability under strongly chelating conditions. SDM-mediated substitution of Asn241 buried in the neutral protease of *B. stearothermophilus* by leucine resulted in an increase in thermostability of $0.7 \pm 0.1°C$ (55). The thermostability of the neutral protease from *B. subtilis* was increased by 0.3 and 1.0°C by replacing Lys with Ser at positions 249 and 290, respectively, whereas the Asp249 and Asp290 mutants exhibited an increased stabilization by 0.6 and 1.2°C, respectively (54).

A protein engineering study was undertaken by Bruinenberg et al. to determine the functions of one of the largest loop insertions (residues 205 to 219), predicted to be spatially close to the substrate-binding region of the SK11 protease from *L. delbrueckii* and susceptible to autoproteolysis (28). Deletion or modification of this loop was shown to affect the activity and autoprocessing of the protease. Graham et al. showed that random mutagenesis of the substrate-binding site of α-lytic protease, a serine protease secreted by the soil bacterium *Lysobacter enzymogenes*, generated enzymes with increased activities and altered primary specifities (77). Substitution of His120 by Ala in the LasA protease of *P. aeruginosa* yielded an enzyme devoid of staphylolytic activity. Thus, His120 was shown to be essential for LasA activity (82).

Fungi

Fungal aspartic proteases are able to cleave substrate with "Lys" in the P1 position. Sequencing and structural comparison suggest that two aspartic acid residues (Asp30 and Asp77) may be responsible for conferring this unique specificity. Lowther et al. engineered the substrate specificity of rhizopuspepsin from *Rhizopus niveus* and demonstrated the role of Asp77 in the hydrolysis of the substrates with lysine in the P-1 position (173).

The primary structure of aspergillopepsin I from *Aspergillus saitoi* ATCC 14332 (now designated *A. phoenicis*) was deduced from the nucleotide sequence of the gene (257). To identify the residue responsible for determining the specificity of aspergillopepsin I toward the basic substrates in the substrate-binding pocket, Asp76 was replaced with a Ser residue by SDM. The striking feature of this mutation was that only the trypsinogen-activating activity of the enzyme was destroyed, suggesting the importance of the Asp76 residue in binding to basic substrates.

To elucidate whether the processing of the pro-region occurs by autoproteolysis or by involving a processing enzyme, Tatsumi et al. changed Ser228 to Ala by SDM (287). *S. cerevisiae* cells harboring a recombinant plasmid with mutant Alp did not secrete active Alp into the culture medium. The yeast cells accumulated a protein of 44 kDa, probably a precursor of Alp (the 34-kDa mature Alp plus the 10-kDa pro-peptide), suggesting that autoproteolytic processing of the pro-region was occurring.

Introduction of a disulfide bond by SDM is known to enhance the thermostability of a cysteine-free enzyme. Aqualysin I, a thermostable subtilisin-type protease from *Thermus aquaticus* YT-1, contains four Cys residues forming two disulfide bonds (149). The primary structure of Alp showed 44% homology to that of aqualysin I, and sites for Cys substitutions to form a disulfide bond were chosen in the Alp based on this homology. Ser69, Gly101, Gly169, and Val200 were replaced by Cys in the mutant Alp. Both Cys69-Cys101 and Cys169-Cys200 mutant Alps were expressed in *S. cerevisiae*, and the enzymes were purified to homogeneity. The Cys169-Cys200 disulfide bond was shown to increase the thermostability as well as the thermoresistance of *Aspergillus oryzae* Alp (110).

In vitro mutation of an aspartic acid residue predicted to be in the active site abolished the barrier activity of *S. cerevisiae* (177). BAR1 possesses a carboxyl-terminal domain of an unknown function, and deletion of 166 of 191 aa of this region had no significant effect on the barrier activity.

Viruses

The protease of HCMV was rendered stable by conversion of one of the three Val141, Val207, or Val254 residues to Gly by SDM (151). The resulting stable proteases are useful as screening tools for HCMV antiviral agents and as diagnostic tools for diseases resulting from HCMV infection.

Replacement of Asp64, a residue from the catalytic core sequence among aspartyl proteases, with Gly was shown to abolish the correct processing of the 53K gag precursor by HTLV-1 gag protease (87).

In poliovirus, the mutation of highly conserved residues, e.g., Cys147 or His161, produced an inactive enzyme while mutation of a nonconserved residue, Cys154, had only a negligible effect on the proteolytic activity (117).

The protein-engineering technique has been exploited successfully for obtaining proteases which show unique specificity and/or stability at high temperature and pH. It has also contributed substantially to our understanding of the structure-function relationship of proteases. In future, protein engineering will offer possibilities of generating proteases possessing entirely new functions.

SEQUENCE HOMOLOGY

Studies of DNA and protein sequence homology are important for a variety of purposes and have therefore become routine in computational molecular biology. They serve as a prelude to phylogenetic analysis of proteins and assist in predicting the secondary structure of DNA and proteins. Proteases are a complex group of enzymes and vary enormously in their physicochemical and catalytic properties. The nucleotide and amino acid sequences of a number of proteases have been determined, and their comparison is useful for elucidating the structure-function relationship (5). The homology of proteases with respect to the nature of the catalytic site has been studied (12, 13). Accordingly, the enzymes have been allocated to evolutionary families and clans. It has been suggested that there may be as many as 60 evolutionary lines of peptidases with separate origins. Some of these contain members with quite diverse peptidase activities, and yet there are some striking examples of convergence (236).

A number of reports on the homology of proteases are available. Takagi et al. found that the thermostable proteases

of *Bacillus stearothermophilus* and *B. thermoproteolyticus* are 85% homologous and the thermolabile proteases of *B. subtilis* and *B. amyloliquefaciens* are 82% homologous, whereas the thermostable protease of *B. stearothermophilus* is only 30% homologous to the thermolabile protease of *B. subtilis* (279). However, an amino acid sequence of 17 residues, which also includes the active-site histidine residue, was found to be highly conserved in all four neutral proteases, suggesting that they have the same three-dimensional structure around the active site despite the difference in their source and physicochemical properties such as thermostability.

Koide et al. compared the amino acid sequences of intracellular serine proteases from *B. subtilis* with those of subtilisin Carlsberg and subtilisin BPN' and showed that they were 45% homologous (138). The sequence around the catalytic triad of serine, aspartate, and histidine is highly conserved, suggesting that the genes for both the intracellular and extracellular proteases have evolved from a common ancestor by divergent evolution (200).

The amino acid sequence of an extracellular alkaline protease, subtilisin J, is highly homologous to that of subtilisin E and shows 69% identity to that of subtilisin Carlsberg, 89% identity to that of subtilisin BPN', and 70% identity to that of subtilisin DY. The amino acid sequence of subtilisin J is completely identical to that of the protease from *B. amylosacchariticus* except for two amino acid substitutions, Thr130 to Ser130 and Thr162 to Ser162, in addition to one amino acid substitution in the signal peptide and two in the propeptide region. The probable active-site residues of subtilisin J, i.e., Asp32, His64, and Ser221, are identical to those of other subtilisins from *Bacillus*. Therefore, it was concluded that the alkaline protease from *B. stearothermophilus* is a subtilisin. Similarly, the various *Bacillus* serine alkaline proteases, such as bacillopeptidase F, subtilisin, Epr, and ISP-1, show considerable homology and conserved amino acids around the active-site residues, i.e., Ser, Asp, and His (265).

The extracellular proteases of *B. subtilis* are synthesized as prepro-enzymes. Four neutral proteases from bacilli with known pro-sequences were compared, and considerable homology within the pro-peptide region was observed (297). Since the pro-peptide region mediates the folding of the protease, it would be interesting to learn about the residues essential for folding and to determine whether the mechanism of folding is similar in these proteases. Sequences corresponding to the mature form of these enzymes were compared by using thermolysin sequence as a reference. The zinc-binding site (His142, His146, and Glu166) and the residues participating in the catalytic reaction and positioning of the substrate backbone in the active site (Asn112, Ala113, Glu143, Tyr157, and His231) were found to be conserved. Differences in these might lead to altered substrate specificities. Of the four calcium-binding sites in thermolysin, two sites, i.e., sites 3 and 4, are absent in the thermolabile neutral proteases of *B. amyloliquefaciens* and *B. subtilis* (NprA) whereas in NprB, Asn187 in site 3 is replaced by Arg. Such changes are responsible for the loss of thermostability and can be detected by sequence homology studies.

Alkaline proteases from various species of *Aspergillus* also show a high degree of homology (131). Alp from *A. oryzae* shows considerable homology (29 to 44%) to the members of the subtilisin family with conserved active-site residues (288). However, Alp exhibits little homology to mammalian serine proteases such as trypsin and chymotrypsin. The deduced structure of the KEX-1 protein, required for the production of the killer toxin of *Kluyveromyces lactis* contains an internal domain with a striking homology to the sequences of subtilisin-

type proteases (242). Therefore it was deduced that the product of the *KEX-1* gene of *K. lactis* is a protease involved in the processing of the toxin precursor.

The characteristic of trypsin-related enzymes is the presence of disulfide bonds, which are absent in all known subtilisins. Proteinase K from *Tritirachium album* Limber is a single chain protein of 277 aa with two disulfide bonds at positions 34-124 and positions 179-248 and a free -SH group at position 73. Sequences around the active-site residues correspond to those around the active-site residues of subtilisins. Comparison of the proteinase K sequence with known subtilisins shows 35% homology and 44% sequence identity to thermitase, which is indicative of a relationship between proteinase K and the subtilisin family. It is likely that these enzymes have evolved from a common ancestral precursor serine proteinase (122). However, there is a distinct difference between the typical subtilisins and proteinase K, since the latter has two disulfide bonds, which are lacking in subtilisins. Therefore, it has been assumed that the two progenitors diverged from an ancestral proteinase, separating the subtilisin-related enzymes into two subclasses: (i) cysteine-containing subtilisins e.g., proteinase K and thermitase, and (ii) cysteine-free subtilisins, e.g., subtilisin Novo, Carlsberg, or DY.

The proteasome or multicatalytic endopeptidase complex is a high-molecular-mass multisubunit complex that is ubiquitous in eukaryotes and also found in the archaebacterium *Thermoplasma acidophilum* (336). While eukaryotic proteasomes contain 15 to 20 different subunits, the archaebacterial proteasome is made of only two different subunits (α and β), yet the complexes are almost identical in size and shape. The α (233-aa) and β (211-aa) subunits of *T. acidophilum* have a sequence identity of 24% and an overall similarity of 47%, indicating that the genes encoding the two subunits arose from a common ancestor. All the sequences of proteasomal subunits from eukaryotes available to date can be related to either the α or β subunit of the *T. acidophilum* "urproteasome," and they can be distinguished by the presence or absence of a highly conserved N-terminal extension which is characteristic of α-type subunits. In terms of evolution, the genes for these α and β subunits can be considered paralogous (genes resulting from duplication and divergence of one gene within one genome) and therefore are able to acquire different functions. The α subunit of the *T. acidophilum* proteasome shows sequence similarity to the *S. cerevisiae* wild-type suppressor gene *scl1*-encoded polypeptide, which is probably identical to the subunit YC7-α of the yeast proteasome. This lends support to a putative role of proteasomes in the regulation of gene expression (337). The amino acid sequence of *Xenopus* proteasome subunit XC3 is highly homologous (95.3%) to those of the rat RC3 and human HC3 subunits (66). The presence of an accessible nuclear targeting signal at the C terminus of the subunits suggests that it is probably involved in the regulation of the cellular distribution of the proteasome.

The secretable acid protease of the yeast *Saccharomycopsis fibuligera* carries a hydrophobic amino-terminal segment of about 20 aa which resembles signal sequences found in a wide variety of secretory protein precursors (95). Alignment of this sequence with the aspartyl protease family showed significant homologies, especially in the regions surrounding the two active-site aspartate residues. These results suggest that the *PEP1* gene is a structural gene for the secretable acid protease of *S. fibuligera*. The aspartic protease from *Rhizopus niveus* (RNAP) shows 76% homology to rhizopuspepsin, 42% homology to penicillopepsin, and 41% homology to human pepsin (100, 101). The homology between RNAP and rhizopuspepsin is found throughout their structures. Based on this homology,

an intron within the coding region and a prepro-enzyme sequence of 66 aa upstream of the mature enzyme were detected in RNAP. Studies of the homology of proteases have shown that the residues involved in the substrate and metal ion binding, catalysis, disulfide bond formation and active-site formation are conserved. Analysis of sequence homology is used in deciphering the structure-function relationship of proteases.

EVOLUTIONARY RELATIONSHIP OF PROTEASES

Proteases are present in all living organisms and are considered to have arisen in the earliest phases of biological evolution, some 1 billion years ago. Comparisons of amino acid sequences, three-dimensional structures, and mechanism of action of proteases assist in deciphering of their course of evolution. Changes in molecular structure have accompanied the demands for altered functions of proteases during evolution. We have compiled the amino acid sequences of proteases from diverse origins such as microbes, plants, and animals and have arranged them in three different groups based on the pH of their action. These sequences, which have been selected from SWISS-PROT and PIR entries, are of comparable length and have been aligned with CLUSTAL W software for multiple alignments (291) (Fig. 4).

Acidic Proteases

The proteases selected here for comparison of amino acid sequences are active between pH 2 and 6. They include mostly aspartic proteases and also some of the cysteine proteases and metalloproteases. They are about 380 to 420 aa long and have different amino acid residues constituting the active site, as shown in Table 6. The homology between these acidic proteases is shown in Fig. 4A. The sequences belonging to pepsin family (A1) are grouped and are aligned below the other sequences. As expected, there is considerable homology among these five acidic proteases. The sequences around the two aspartic residues (D97 and D258, residues numbered according to the Bajra protease) constituting the active site are conserved. Among these five proteases, the rat and monkey proteases show maximum homology (68.4%) and are related to the mosquito lysosomal aspartic protease. When four monkey pepsinogens which show development-dependent expression were compared, a very high homology was observed (126). Pepsinogens A-1 and A-2/3 differed in seven amino acids and only in five amino acids when the pepsin moiety alone was examined. The mosquito lysosomal protease is very closely related to human cathepsin D, exhibiting 92% homology (37).

The amino acid sequences of C. tropicalis and Saccharomycopsis fibuligera show considerable homology (42.6%). High similarity scores were obtained when the acid protease from C. tropicalis was compared with Rhizopus aspartic proteases, human pepsinogen A precursor, protease A from yeast, the barrier protein from S. cerevisiae, and an acidic protease from S. fibuligera (294).

The cysteine protease from Hordeum vulgare shows some homology to the snake venom metalloprotease from Crotalus atrox, which is not statistically significant, whereas the Gpr protease from Bacillus megaterium, which plays a vital role in spore germination, shows least homology to all other acidic proteases but shares one of the active-site aspartate residues (D258) with them. The Gpr acidic proteases of B. subtilis and B. megaterium showed 68% identity in their

sequences, but comparison of the B. subtilis Gpr amino acid sequence with that of its serine protease or metalloprotease revealed no significant homologies (274), which supports our observations. This suggests that the genes encoding these proteases do not have a common ancestor or that if they do so, they have undergone much divergence. The lack of homology between the spore protease and other B. subtilis proteases can be explained by differences in their properties such as the number of subunits and sequence specificity for the substrate. Thus, our results, in agreement with previous reports, indicate that the extent of homology is greater if the proteases belong to the same family and that in the same family the homology is greater if the phylogenetic distance is shorter.

A pairwise computer comparison also provides more information about the evolutionary relationships between the members of the different families. The dendrograms generated by this analysis, using the TreeView package (213), demonstrate the relationship among the proteins based on the similarity of the amino acid sequences (Fig. 5a).

Neutral Proteases

The neutral proteases, which are active at neutral or weakly alkaline or weakly acidic pH, include cysteine proteases, metalloproteases, and some of the serine proteases. Brenner (25) has pointed out that the two codons for serine TCN and AGY cannot be interconverted by single nucleotide mutations but can be connected by two other codons, ACN for threonine and TGY for cysteine. Thus, there can be at least two different lines of descent for the active-site sequences of the serine proteases. The simplest pathway for this convergent evolution is by the divergence of each line from a precursor which was itself catalytically active and had much the same sequence. It was further demonstrated that modern serine enzymes are likely to have arisen from cysteine precursors. These findings encourage the search for evidence to connect the presumed and existing cysteine sequences with their postulated metalloenzyme predecessors. For this search and construction of phylogenetic trees, gene structure is important. Thus, multiple lines of descent can be realistically considered in situations with sequence similarity but with differences in gene structure.

The neutral proteases selected for sequence analysis in the present study are in the size range of 225 to 275 aa (Table 6). The homology between them is shown in Fig. 4B. Of 14 proteases, 9 belong to the T1A or proteasome A family of the multicatalytic endopeptidase complex. The sequences of the proteasomal subunits aligned here can be related to the α subunit of the Thermoplasma proteasome and show considerable homology. It is still not clear to which family of proteases the proteasomes belong (93). As in the cysteine family of proteases, all nine proteasome subunits show a conserved proline residue (P-17), which may serve to prevent unwanted N-terminal proteolysis (12). Many residues at the N terminus are highly conserved, which is a characteristic of the α-type subunits. The similarity decreases toward the C terminus which appears to be rather variable (337). Although the β subunit shows no sequence motif characteristic of serine proteases, it contains all the essential amino acid residues forming the catalytic triad or the "charge relay system" (Ser, Asp, and His). These residues are found to be conserved (Ser16, His73, and Asp84), except for the histidine in the α subunits of Thermoplasma, yeast (S. cerevisiae), and Caenorhabditis elegans (residues are numbered according to the Thermoplasma α-subunit

A

```
CYS2_HORVU   MGLLSKKLLVASMVAAVLAYAAYEL-CSAIPMEDKDLESEEALVDLYERMQSAHRYRRHH 51
HRTD_CROAT   -MIEVLLVTICLAVFPYQGSSIILESGNVNDYEYVYTPRKVTALPKGAYQPKYEDAHQYEL
GPR_BACME    --MEKELDLSQYSVRTDLAVEAKD--IALENQPKPHNQSE--IKGVIYKEKEEQGYKISM
ASPP_AEDAE   NLIKSITALYCLAVLSQADFYRVQ--LHKTESARQHFRM---VDTEIKQLRLKYN-----
CARP_CANTR   --MATIFLFTKHVFIALAFALFAQG-LTIPDGIEKRTDKVYSLDFTVIRKPFNATAHRLI
CARP_SACFI   --MLFSKSLLLSVLASLSFAAPYE--KREKTLTLDFDVKRISSKAKNYTVASSPGFRRNL
PEPC_RAT     --NKVNVYALLCLPLLEASLLRVP--LRKMKSIRETKKEQGVLKDFLKTHKYDPGQKYHF
PEP2_MACFU   --NKVLLLLGLVA-LSECIIHKVP--LVRKKSLRRHLSEHGLLKDFLKKHNFNPASKYFP
                 ●                                         ●
```

```
CYS2_HORVU   A--EKHRRFGTFKS--NAHFIHSHHKRGDHPYRLHLMRFGDMDQAEFRAT--FYGDLR-- 111
HRTD_CROAT   KVHGEPVVLHLEKNKELFSKDYSETHYS-PDGRKITTMP-SVEDHCYYRG--RIENDA--
GPR_BACME    Y--EITEEGAEAIG-KKKGRYVTLESVG-IREQD-TEKQ-EEAMEEVFAK---ELNFF--
ASPP_AEDAE   ---AYSGPVPEPLSNYLDAQYYGAITIG-ITPPQSFKVYF-DTGSSNLWVP--SKECSFTM
CARP_CANTR   Q--KRSDYPTTLIN--EGPSYAADIVVG-ISNQQKQTYVI-DTGSSDLWVVDTDAECQVTY
CARP_SACFI   RAASDAGYTISLEN--EYSFYLTTIEIG-ITPGQKLQVDV-DTGSSDLWVP--GQGTSS--
PEPC_RAT     GNFGDYSVLYEPNA-YNDASYFGEISIG-ITPPQNFLVLF-DTGSSNLWVS--SYVYCQS--
PEP2_MACFU   QAEAPTLIDEQPLENYLQMEYFGTIGIG-ITPAQDFTVIF-DTGSSNLWVP--SYYCSS--
                                   ●              *          ●●
```

```
CYS2_HORVU   ---------RDTPSKPPSVPGFMY--AALMVSDLPPSYDVRQKGAVTGVKDQGKCGSCM 159
HRTD_CROAT   ----------DSTASISACNGLKG---HFKLQGEMYL-IEPLELSDSEAHA--VFKLENV
GPR_BACME    ----------IKSLNIPDDASCLVVGLGMLSVTPD----ALGPKAVDNLLITRHLFELQPE
ASPP_AEDAE   I-----ACLMHNKYNAKKSSTFEKNGTAFHIQYGSGS-LSGYLSTDTVGLG--GVSVTKQ
CARP_CANTR   SGQTNNFCKQEGTFDPSSSSSAQNLNQDFSIEYGDLTSSQGSFYKDTVGFG--GISIKNQ
CARP_SACFI   ---------LYGTYDHTKSTSYKKORSGFSISYGDGSSARGDVAQETVSIG--GASIITGL
PEPC_RAT     E-----ACTTHARFNPSKSSTYYTEGQTFSLQYGTGS-LTGFFGYDTLTVQ--SIQVPNQ
PEP2_MACFU   L-----ACTMHNRFNPQDSSTYQSTSGTVSITYGTGS-MTGILGYDTVQVG--GISDTNQ
                                     ●              ●       ●
```

```
CYS2_HORVU   AFSTYVSVE-GDNAIRTGSLYS-LSEQELIDCDTAD----NDGCQGGLNDNAFEYIKNH- 212
HRTD_CROAT   EKEDEAPKNCGVTQNVESYEPIKKASDLNLKPDQQN-LPQRYIELYVVADHRYFNKYNSD
GPR_BACME    SYQDG-FRP--VSAIYPGVNGNTGIETSDIIFGYVK----KVNPDFIIAIDALAARSIER
ASPP_AEDAE   TFAEAINEP-GLVFYAAKFDGILGLGYSSISVDGVVPYFYNNFNQGLIDAPVFSFYLNRD
CARP_CANTR   QFADYTTTSVDQGINGIGFIAY-EAGYNLYSNYPYT-----LKKQGIINKNAYSCDLNSE
CARP_SACFI   EFGDATSQDYGQGLLGIGLKGKEASAQSSNSFTYDN-LPLKLKDQGLIDKAAYSLYLNSE
PEPC_RAT     EFGLSENEP-GTNFYYAQFDGINGLAYPGLSSGGATTALQGMLGEGALSQPLFGVYLGSQ
PEP2_MACFU   IFGLSETEP-GSFLYYAPFDGILGLAYPSISSSGATPYFDNIVNQGLVSQDLFSVYLSAD
                                                          ●  ●●
```

FIG. 4.

```
CYS2_HORVU   -----GGLITEAAYPYRA-ARGTCM---VARAAQHSPVYYHID----GHQDYPAH-----  254
HRTD_CROAT   LHT--IRTRYHEIVHFIHGFYRSLH----IHVSLTDLEIVSHED-QIHIQSASSDT----
GPR_BACME    -----VHATIQIS-DSGI-HPGSGV---GHKRKEISYETLGIP.VIAIGIPTYVDAVS---
ASPP_AEDAE   PSAAEGGEIIFGGSDSMK-YTGDFT---YLSVDRKAYWQFKHDSYKYGDTEFCMH-----
CARP_CAHTR   DAS--TGKIIFGGVDHAK-YTGTLT---ALPVTSSVELRYHLGSIHFDGTSYSTH-----
CARP_SACFI   DAT--SGSILFGGSDSSK-YSGSLATLDLYHIDDEGDSTSGAYAFFYELEGIEAGSSSIT
PEPC_RAT     QGSH-GGQIVFGGVDKHL-YTGEIT---HVPVTQELYWQITIDDFLIGDQASGWCSS---
PEP2_MACFU   DQS--GSYVIFGGIDSSY-YTGSLH---VVPVSYEGYWQISYDSITHHGEAIACA-----
```

```
CYS2_HORVU   ---SE--EDLARAVAHQPYSVAVEASGKAFHF-YSEG--YFTGECGTELD------HGVA  300
HRTD_CROAT   --LHAFAEVRETDLLHRKSHDHAQLLTAIELDEETLGLAPLGTHCDPKLSIGIVQDHSPI
GPR_BACME    --ITSDTIDFILKHFGREHKEQGKPSKSLLPSGHTFGE-KKKLTEDDLPH------EEQR
ASPP_AEDAE   --GCEAIADTGTSLIAGPYSEVTAIHKAIGGTPIMHG--EYHVDCSLIPKL-----PKIS
CARP_CAHTR   ---ADVVLDSGTTITYFSQSTADKFARIYGATHDSRHEIYRLPSCDLSG------DAVV
CARP_SACFI   KTTYPALLDSGTTLIYAPSSIASSIGREYGTYSYSYG--GYYTSCDATG------PDFK
PEPC_RAT     -QGCQGIVDTGTSLLVHPAQYLSELLQTIGAQEGEYG--EYFVSCDSVSSL-----PTLS
PEP2_MACFU   -EGCQAIVDTGTSLLTGPTSPIAHIQSDIGASEHSDG--EHYVSCSAISSL-----PDIY
```

```
CYS2_HORVU   YVGYGVAEDGK-AYWTYKHSVG-PSWGEQGYIRVE---KOS-GASGGLCGI------AHE  348
HRTD_CROAT   HLLHGVTHAHELGHHLGHEHDG--KDCLRGASLCIHRPGLTKGRSYEFSDDSHHYYERFL
GPR_BACME    QTYLGHIGTLP-DEEKRRLIHE--VLAPLGHKLHVT-PKEY-DHFIEDHAH-----VVA
ASPP_AEDAE   FVLGGKSFDLEGADYVLRYAQHGKTICLSGFHGIDI-PPPH-GP-LHILGD-----VFI
CARP_CAHTR   HFDQGVKITVPLSELILKDSDS--SICYFGISRHD-----------AHILGD-----KFL
CARP_SACFI   FSFHGKTITVPFSHLLFQHSEG-DSECLVGYLSS--------GSHYYILGD-----AFL
PEPC_RAT     FVLHGVQFPLSPSSYIIQEDHF----CHVGLESISL-TSES-GQPLWILGD-----YFL
PEP2_MACFU   FTIHGIQYPVPPSAYILQSQGS----CISGFQGHDY-PTES-GE-LHILGD-----YFI
```

```
CYS2_HORVU   ASYP-YKTYSKPKPTPRRA--LGARESL--------  373
HRTD_CROAT   KQYKPQCILHKPLRIDPVSTPVSGHELLEAGEE---
GPR_BACME    GGL--HAALHHE--VDQEH--FGAYTH---------
ASPP_AEDAE   GKY--YTEFDHG--HORYG--FATAV----------
CARP_CAHTR   RRA--YIVYDLD--DKTIS--LAQVKYTSSSDISAL
CARP_SACFI   RSA--YVYYDIO--HSQVG--IAQAKY---------
PEPC_RAT     RSY--YAIFDHG--HHKVG--LATSV----------
PEP2_MACFU   RQY--FTVFDRA--HHQYG--LAPVA----------
```

FIG. 4—Continued.

B

PRCA_THEAC	M Q Q G Q K A Y D R A I T V F S - - - P - - D G R L F Q Y E - - - Y A R E A Y K K G S T A L G K K F A H G - - V L L I S 50
PRC3_YEAST	- - M T D R - Y S F S L T T F S - - - P - - S G K L G Q I D - - - Y A L T A Y K Q G Y T S L G I K A T H G - - V Y I A T
PRC6_SCHPO	- - - M S G - Y D R A L S V F S - - - P - - D G R L L Q Y E - - - Y G Q E A Y R R G T T A I A L R G K E C - - I Y I G V
PRC2_ORYSA	- - M F R K Q Y D T D V T T V S - - - P - - A G R L F Q Y E - - - Y A M E A Y K Q G S A C V G L R S R T H - - A Y L A A
PRC6_ARATH	- - - M A R - Y D R A I T V F S - - - P - - D G H L F Q Y E - - - Y A L E A V R K G H A A V G V R G T D T - - V Y L A V
PRC8_DICDI	- M S S A R - Y D R A I T V F S - - - P - - D G H L F Q Y E - - - Y A M E A V R K G T V A V G Y R G K D V - - I Y L G V
PRC8_CAEEL	- S S I G T G Y D L A A S T F S - - - P - - D G R I F Q Y E - - - Y A Q K A V D K A G T M I A I R G K H G - - V Y V V A
PRC6_DROME	- - M S S R - Y D R A V T I I F S - - - P - - D G H L L Q Y E - - - Y A Q E A Y R K G S T A V G Y R G A N C - - V Y L G V
PRC3_XENLA	- - - A E R G Y S F S L T T F S - - - P - - S G K L V Q I E - - - Y A L A A Y A A G A P S V G I K A T H G - - V Y L A T
CAHS_BOVIN	- K F L V K S F L K G G G G G G G G G - L G G G L G K V L G G L I S G A G G G G G G G G G G G G G G - - T A K R I
CAHS_PIG	- K F L V K S F L K G G G G G G G G G G L G G G L G K V L G G L I S G A G G G G G G G G G G G G G G G G T A K R I
TRYT_CANFA	- - - K P S P L Y L A L A L L G S L V P - V S P A P G Q A L - - - - Q R V G I V G G R E A P G S K V P V Q - - V S L R L
TRYB_HUMAN	- - - M L K L L L L A L P V L A S R A Y - A A P A P G Q A L - - - - Q R V G I V G G Q E A P R S K V P V Q - - V S L R Y
SHPA_STRLI	- - - K R I T L P L L S T A V G - - - - - - L G L T A A V L - - - G T G P A A T A A A P Q E P V R A A Q - - - - - L G Y

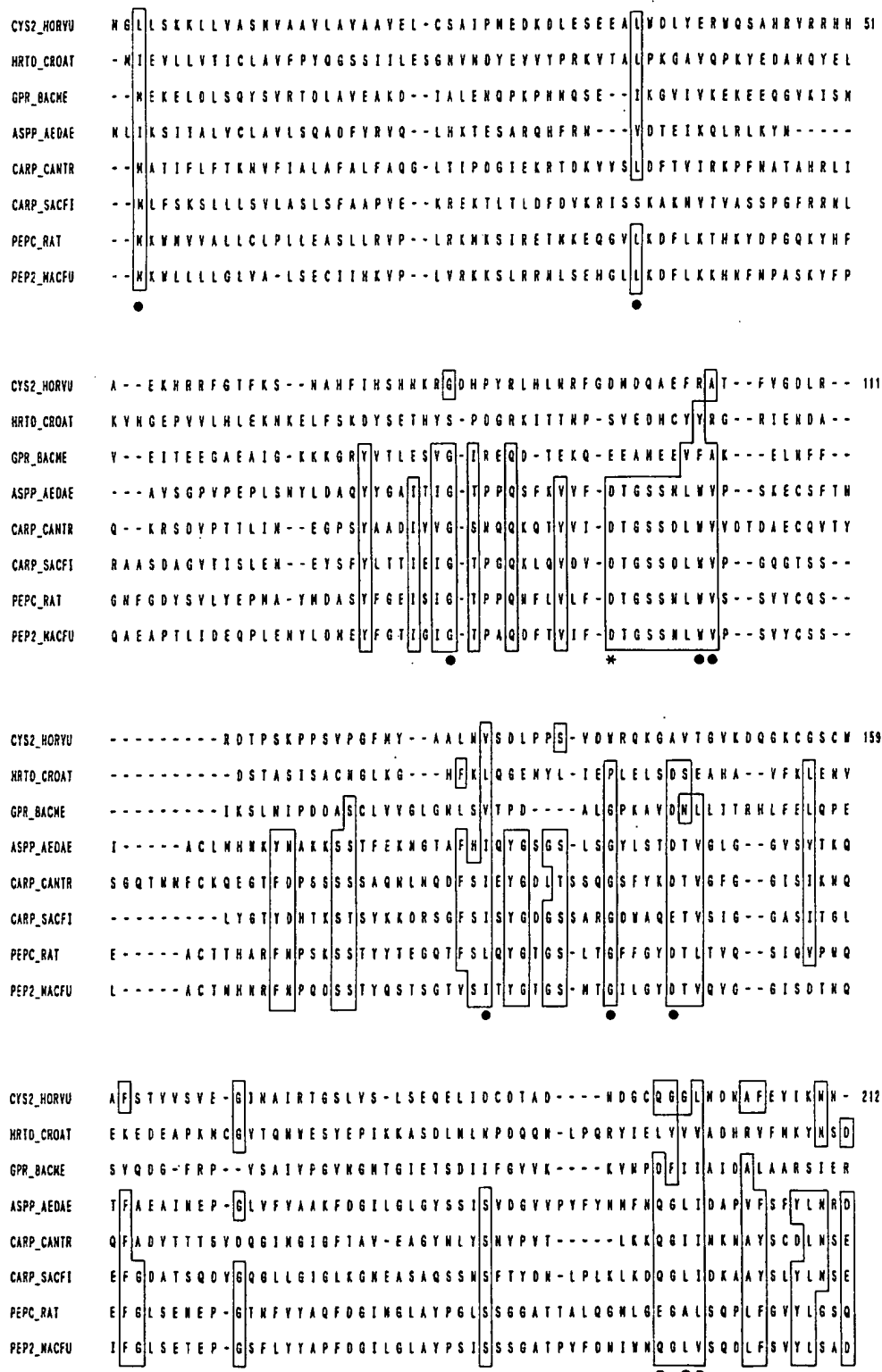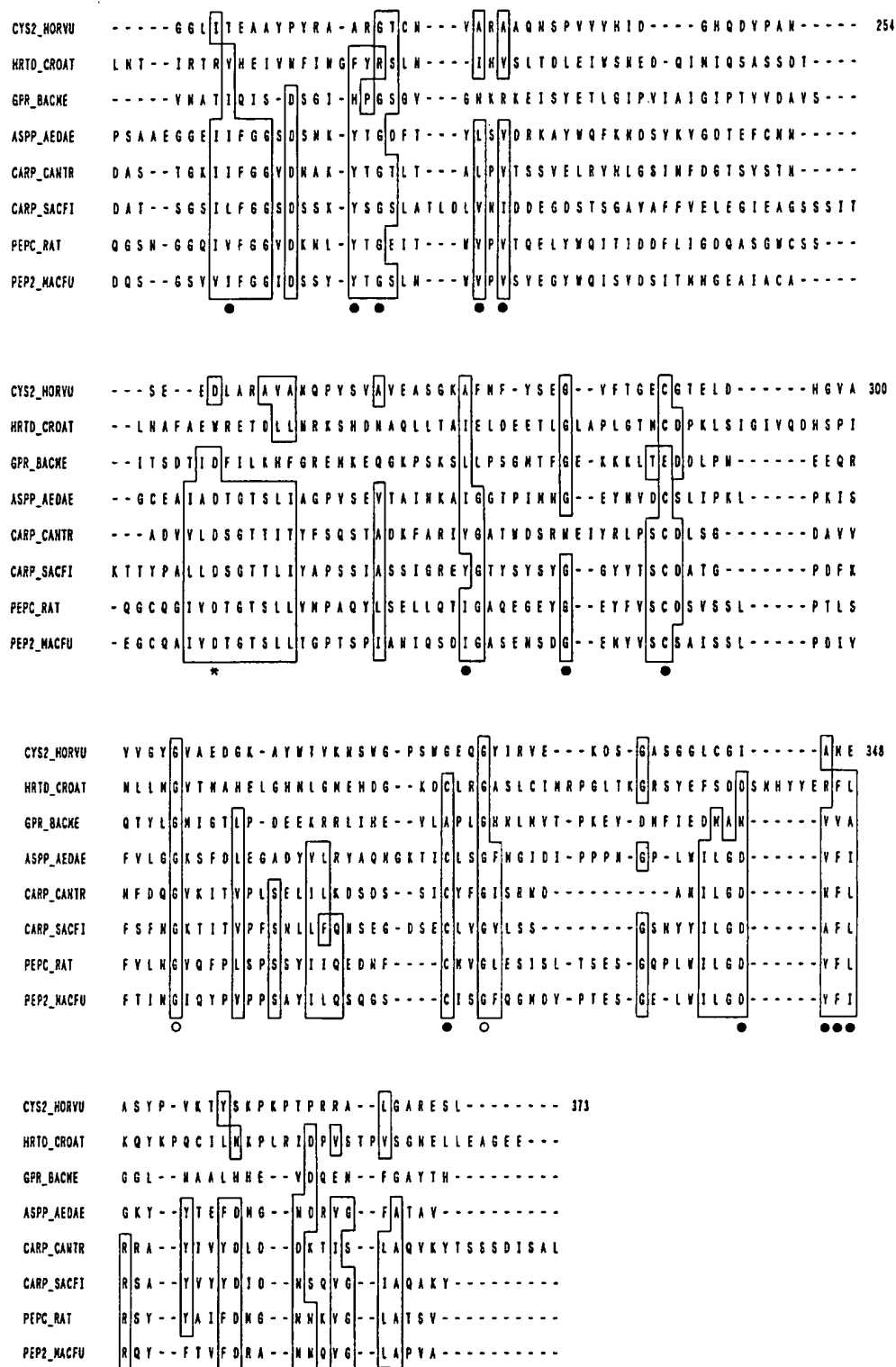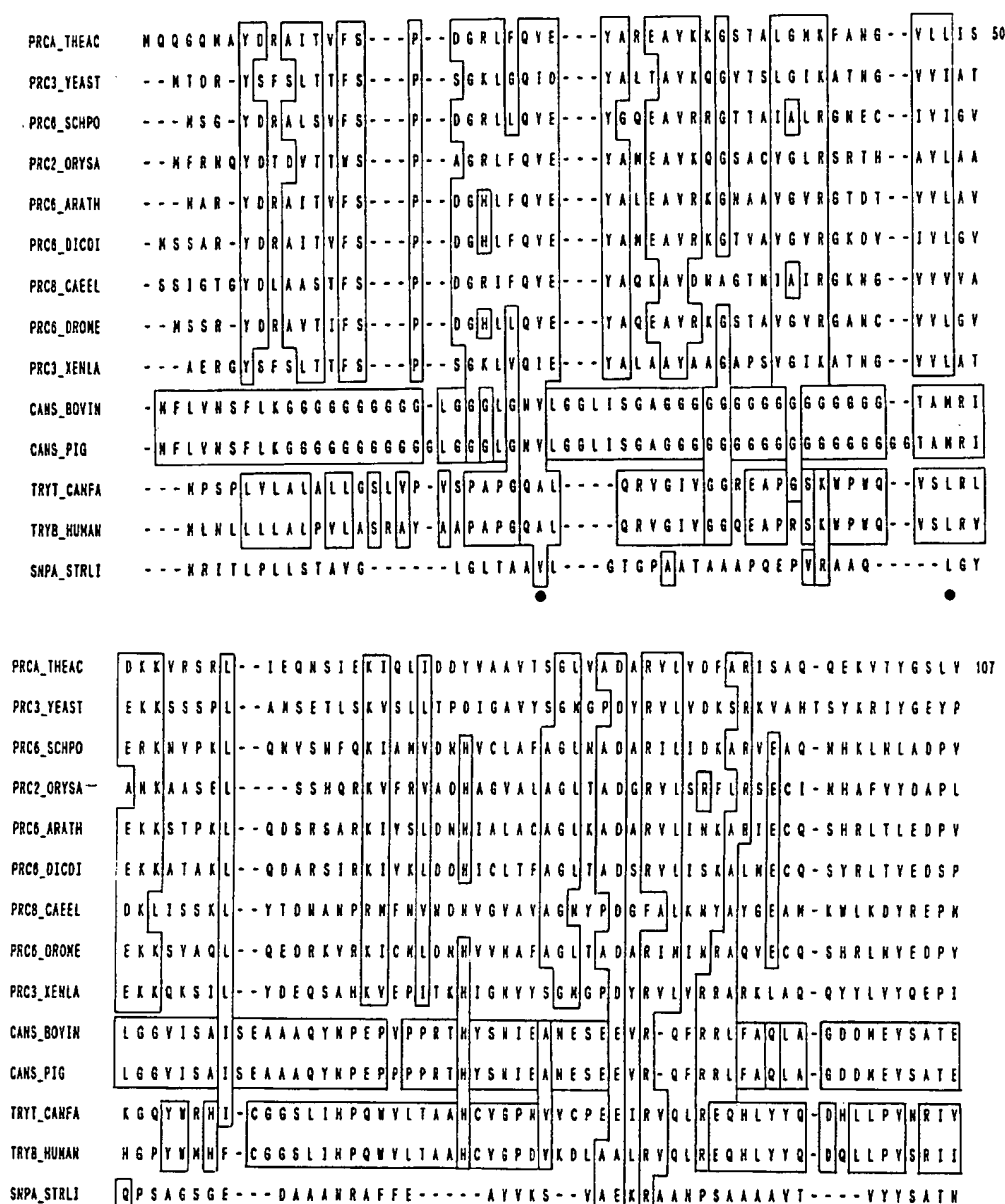|	
PRCA_THEAC	D K K V R S R L - - I E Q K S I E K I Q L I D D Y V A A V T S G L V A D A R V L V D F A R I S A Q - Q E K V T Y G S L Y 107
PRC3_YEAST	E K K S S S P L - - A K S E T L S K V S L L T P D I G A V Y S G K G P D Y R V L V D K S R K V A H T S Y K R I Y G E Y P
PRC6_SCHPO	E R K K V P K L - - Q K V S K F Q K I A K V D K H V C L A F A G L K A D A R I L I D K A R V E A Q - K H K L K L A D P V
PRC2_ORYSA	A K K A A S E L - - - - - S S H Q R K V F R V A D H A G V A L A G L T A D G R V L S R F L R S E C I - K H A F V Y D A P L
PRC6_ARATH	E K K S T P K L - - Q D S R S A R K I V S L D K H I A L A C A G L K A D A R V L I K K A B I E C Q - S H R L T L E D P V
PRC6_DICDI	E K K A T A K L - - Q D A R S I R K I V K L D D H I C L T F A G L T A D S R V I I S K A L H E C Q - S Y R L T V E D S P
PRC8_CAEEL	D K L I S S K L - - Y T D M A K P R K F K V K D K V G V A V A G K Y P D G F A L K K Y A V G E A K - K M L K D Y R E P K
PRC6_DROME	E K K S Y A Q L - - Q E D R K V R K I C K L D K H V V K A F A G L T A D A R I K I K R A Q V E C Q - S H R L K Y E D P V
PRC3_XENLA	E K K Q K S I L - - Y D E Q S A H K V E P L T K H I G K V Y S G K G P D Y R V L V R R A R K L A Q - Q Y Y L V Y Q E P I
CAHS_BOVIN	L G G V I S A I S E A A A Q Y K K P E P V P P R T H Y S K I E A K E S E E V R - Q F R R L F A Q L A - G D D K E Y S A T E
CAHS_PIG	L G G V I S A I S E A A A Q Y K K P E P P P R T H Y S K I E A K E S E E V R - Q F R R L F A Q L A - G D D K E Y S A T E
TRYT_CANFA	K G Q Y V R H I - C G G S L I H P Q K V L T A A H C V G P H V V C P E E I R V Q L R E Q H L Y Y Q - D H L L P V K R I V
TRYB_HUMAN	H G P Y K K H F - C G G S L I H P Q K V L T A A H C V G P D V K D L A A L R V Q L R E Q H L Y Y Q - D Q L L P V S R I I
SHPA_STRLI	Q P S A G S G E - - - D A A A K R A F F E - - - - - A V V K S - - V A E K R A A H P S A A A A V T - - - - V Y Y S A T K

FIG. 4—*Continued*.

sequence). Therefore, it is possible that the active site is shared between the α and β subunits (336). The tyrosine autophosphorylation site at Tyr123 is conserved in six of the nine sequences. The cAMP/cGMP-dependent phosphorylation sites between aa 31 and 37 are found only in *Thermoplasma* and *Drosophila* spp. (84), as reported by Zwickl et al. (337). A consensus nuclear localization signal between aa 49 and 56 (240) and a region complementary to the nuclear localization signal consensus sequence (326) between aa 201 and 212 can be identified in these sequences. Thus, the sequence comparison of various α proteasome subunits from archaebacteria to mammals shows high homology.

The bovine and porcine proteases which belong to the calpain or C2 family of cysteine proteases differ from each other

in only six amino acid residues and thus show almost 99% homology to highly conserved calcium-binding domains and the N-terminal glycine-rich hydrophobic region. The region rich in proline residues (aa 76 to 81, numbered as in the *Thermoplasma* protease) is also conserved except at position 79, where proline is replaced by valine.

Tryptase precursors from humans and dogs (301), which belong to the S1 or trypsin family of serine proteases, show 76% sequence identity. The signal sequence from residues 1 to 30 is 60% identical; the difference is only in the site of glycosylation, which is Asn132 in the canine sequence and Asn233 in the human sequence. The sequences for active-site and disulfide bond formation are highly conserved and correspond to those of chymotrypsinogen (302).
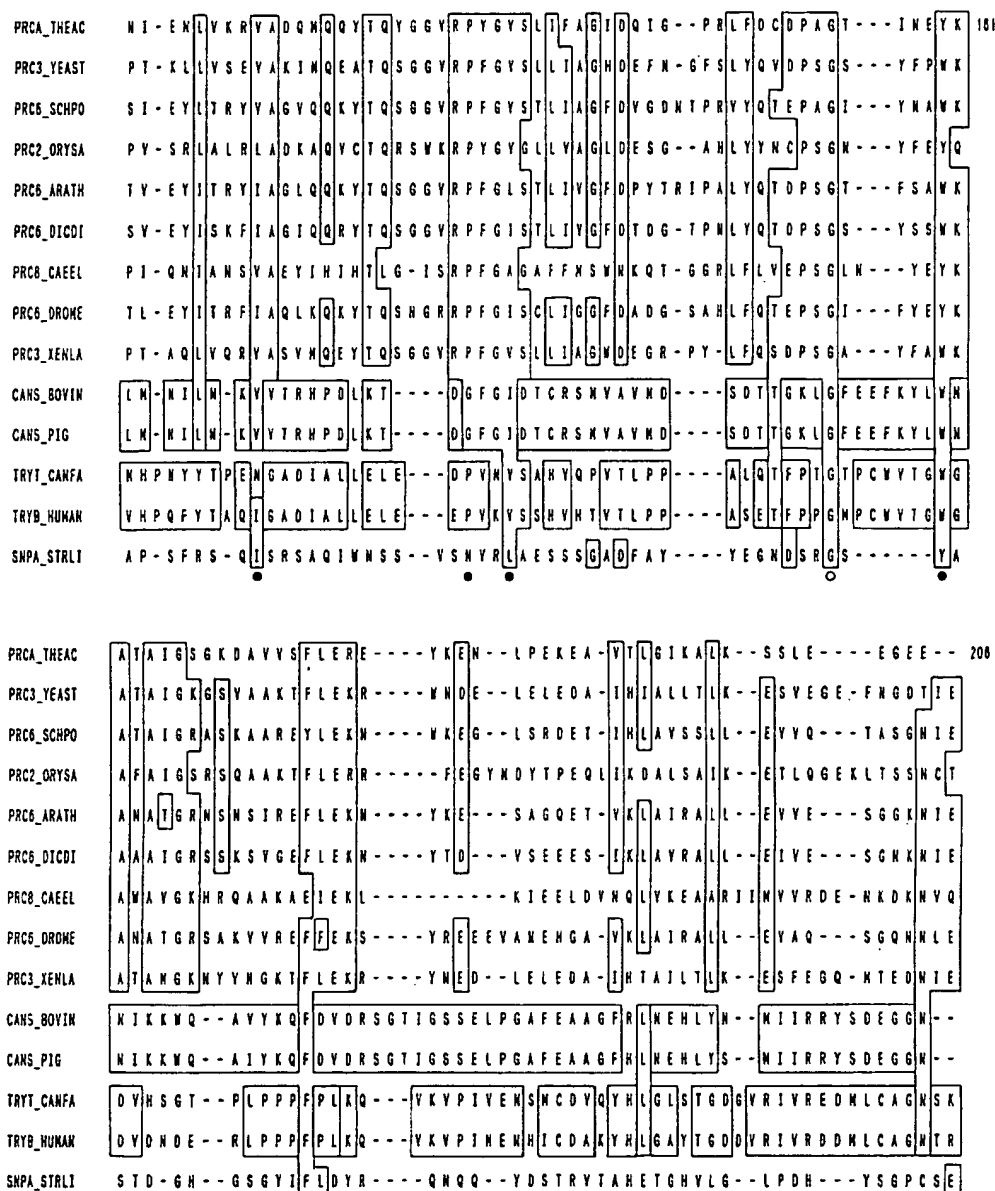
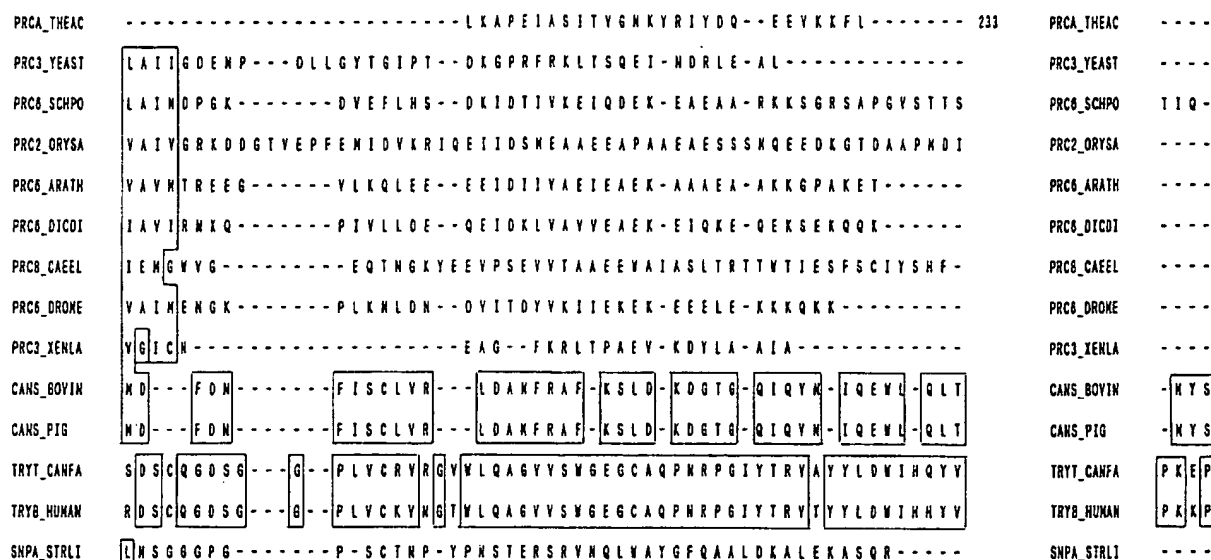PRCA_THEAC   N I - E N L V K R V A D Q M Q Q Y T Q Y G G V R P Y G Y S L I F A G I D Q I G - - P R L F D C D P A G T - - - I N E Y K   151

PRC3_YEAST   P T - K L L V S E V A K I M Q E A T Q S G G V R P F G Y S L L T A G H D E F N - G F S L Y Q V D P S G S - - - Y F P V K

PRC6_SCHPO   S I - E Y L T R Y V A G V Q Q K Y T Q S G G V R P F G Y S T L I A G F D V G D N T P R Y Y Q T E P A G I - - - Y N A V K

PRC2_ORYSA   P V - S R L A L R L A D K A Q V C T Q R S V K R P Y G Y G L L V A G L D E S G - - A H L Y N C P S G N - - - Y F E Y Q

PRC6_ARATH   T V - E Y I T R Y I A G L Q Q K Y T Q S G G V R P F G L S T L I V G F D P Y T R I P A L Y Q T D P S G T - - - F S A V K

PRC6_DICDI   S V - E Y I S K F I A G I Q Q R Y T Q S G G V R P F G I S T L I V G F D T O G - T P N L Y Q T O P S G S - - - Y S S V K

PRC8_CAEEL   P I - Q N I A N S V A E Y I H I H T L G - I S R P F G A G A F F N S V H K Q T - G G R L F L V E P S G L N - - - Y E Y K

PRC6_DROME   T L - E Y I T R F I A Q L K Q K Y T Q S H G R R P F G I S C L I G G F D A D G - S A H L F Q T E P S G I - - - F Y E Y K

PRC3_XENLA   P T - A Q L V Q R V A S V H Q E Y T Q S G G V R P F G V S L L I A G M D E G R - P Y - L F Q S D P S G A - - - Y F A V K

CANS_BOVIN   L N - N I L N - K V V Y T R H P D L K T - - - - D G F G I D T C R S N V A V N D - - - - S D T T G K L G F E E F K Y L V N

CANS_PIG   L N - N I L N - K V V Y T R H P D L K T - - - - D G F G I D T C R S N V A V N D - - - - S D T T G K L G F E E F K Y L V N

TRYT_CANFA   N H P N Y Y T P E N G A D I A L L E L E - - - D P V N Y S A H V Q P V T L P P - - - - A L Q T F P T G T P C V V T G V G

TRYB_HUMAN   V H P Q F Y T A Q I G A D I A L L E L E - - - E P V K V S S H V H T V T L P P - - - - A S E T F P P G N P C V V T G V G

SHPA_STRLI   A P - S F R S - Q I S R S A Q I V N S S - - V S N Y R L A E S S S G A D F A Y - - - - Y E G N D S R G S - - - - - - Y A

PRCA_THEAC   A T A I G S G K D A V V S F L E R E - - - - Y K E N - - L P E K E A - V T L G I K A L K - - S S L E - - - - E G E E - -   201

PRC3_YEAST   A T A I G K G S V A A K T F L E K R - - - - - V N D E - - L E L E D A - I H I A L L T L K - - E S V E G E - F N G D T I E

PRC6_SCHPO   A T A I G R A S K A A R E Y L E K N - - - - - V K E G - - L S R D E T - I H L A V S S L L - - E V V Q - - - T A S G N I E

PRC2_ORYSA   A F A I G S R S Q A A K T F L E R R - - - - - - F E G Y N D Y T P E Q L I K D A L S A I K - - E T L Q G E K L T S S N C T

PRC6_ARATH   A N A T G R N S N S I R E F L E K N - - - - Y K E - - - - S A G Q E T - V K L A I R A L L - - E V V E - - - S G G K N I E

PRC6_DICDI   A A A I G R S S K S V G E F L E K N - - - - Y T D - - - V S E E E S - I K L A Y R A L L - - E I V E - - - S G N K N I E

PRC8_CAEEL   A V A V G K H R Q A A K A E I E K L - - - - - - - - - - K I E E L D V N Q L Y K E A A R I I N V Y R D E - N K D K N V Q

PRC6_DROME   A N A T G R S A K Y V R E F F E K S - - - - Y R E E E V A N E H G A - V K L A I R A L L - - E V A Q - - - S G Q N N L E

PRC3_XENLA   A T A N G K N Y Y N G K T F L E K R - - - - Y N E D - - L E L E D A - I H T A I L T L K - - E S F E G Q - N T E D N I E

CANS_BOVIN   K I K K N Q - - A V Y K Q F D V D R S G T I G S S E L P G A F E A A G F R L N E H L Y N - - N I I R R Y S D E G G N - -

CANS_PIG   N I K K N Q - - A I Y K Q F D V D R S G T I G S S E L P G A F E A A G F H L N E H L Y S - - N I I R R Y S D E G G N - -

TRYT_CANFA   D V H S G T - - P L P P P F P L K Q - - - V K V P I V E N S N C D V Q Y H L G L S T G D G V R I V R E D N L C A G N S K

TRYB_HUMAN   D V D N D E - - R L P P P F P L K Q - - - V K V P I N E N H I C D A K Y H L G A Y T G D D V R I V R D D N L C A G N T R

SHPA_STRLI   S T D - G H - - G S G Y I F L D Y R - - - - Q N Q Q - - Y D S T R V T A N E T G H V L G - - L P D N - - - Y S G P C S E

FIG. 4—Continued.

Alkaline Proteases

The alkaline proteases selected here are active in the pH range of 8 to 13 and are about 420 to 480 residues in length. Six of them belong to the S8 or subtilase family of serine proteases (Table 6). They are aligned in their phylogenetic order, as shown in Fig. 4C. Considerable homology within the same genus is observed for *Bacillus* and *Aspergillus* proteases and three other fungal proteases. However, these proteases show comparatively lower homology among themselves. The active-site residues, as well as the residues around the active site, are highly conserved, suggesting that they may have evolved from a common ancestor. The sequences of *E. coli* and *Cyprinus* carpio seem to be homologous to some extent, but they do not have common active-site residues and therefore do not have a common ancestor. These two, in turn, show no significant homology to the other seven alkaline proteases. The overall homology among all these sequences can be represented by the dendrogram in Fig. 5c.

The results of our analysis of the amino acid sequences of the acidic, neutral, and alkaline proteases indicate that the members of the pepsin family of acidic proteases may have evolved from a common ancestor by convergent evolution. High homology between the sequences of the α subunits of proteasomes provides evidence for the presence of an evolutionarily conserved gene family. No amino acid residues are conserved in all the acidic or neutral proteases, except glycine. The alkaline serine proteases seem to have evolved from a
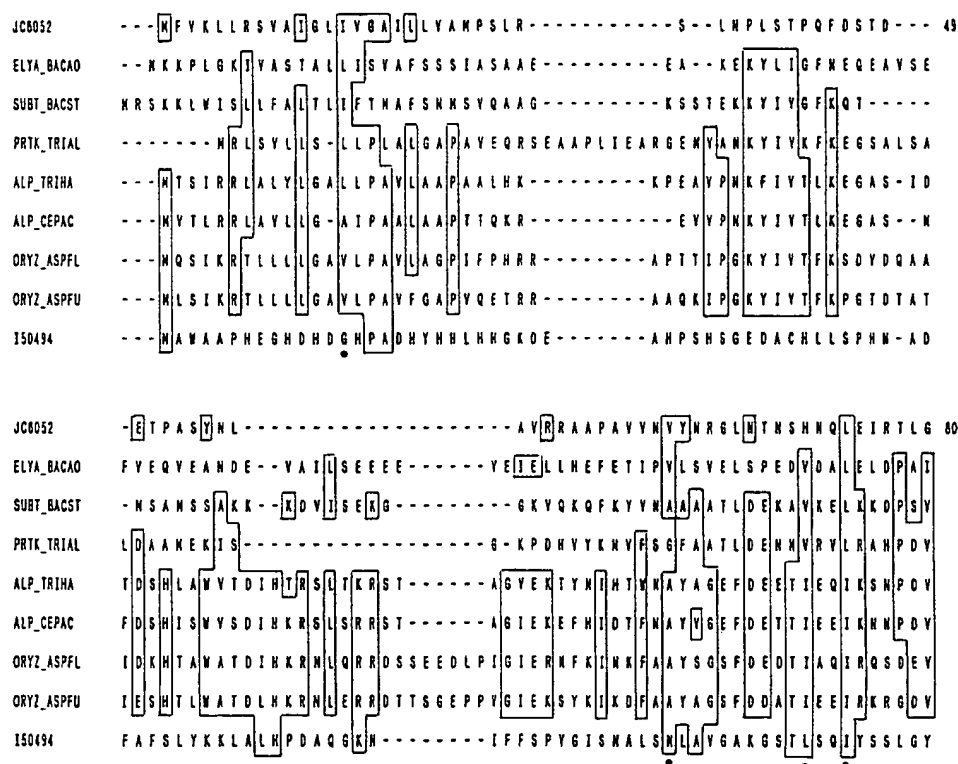
The relationship between these neutral proteases is evident from the dendrogram shown in Fig. 5b.

```
PRCA_THEAC    ------------------------------LKAPEIASITYGNKYRIYDQ--EEYKKFL------- 233      PRCA_THEAC    - - - -
PRC3_YEAST    LAII GDENP---DLLGYTGIPT--DKGPRFRKLTSQEI-NDRLE-AL------------            PRC3_YEAST    - - - -
PRC6_SCHPO    LAIN DPGK-------DVEFLHS--DKIDTIVKEIQDEK-EAEAA-RKKSGRSAPGYSTTS          PRC6_SCHPO    T I Q -
PRC2_ORYSA    VAIY GRKDDGTVEPFENIDVKRIQEIIDSNEAAEEAPAAEAESSSKQEEDKGTDAAPKDI         PRC2_ORYSA    - - - -
PRC6_ARATH    VAVH TREEG------VLKQLEE--EEIDIIYAEIEAEK-AAAEA-AKKGPAKET------          PRC6_ARATH    - - - -
PRC6_DICDI    IAVI RNKQ-------PIVLLDE--QEIOKLVAVYEAEK-EIQKE-QEKSEKQQK------          PRC6_DICDI    - - - -
PRC8_CAEEL    IEH GVYG---------EQTNGKYEEVPSEVVTAAEEVAIASLTRTTVTIESFSCIYSHF-         PRC8_CAEEL    - - - -
PRC6_DROME    VAIN ENGK-------PLKNLDN--DYITDYVKIIEKEK-EEELE-KKKQKK--------          PRC8_DROME    - - - -
PRC3_XENLA    V GIC N-------------------EAG--FKRLTPAEY-KDYLA-AIA----------           PRC3_XENLA    - - - -
CANS_BOVIN    N D ---FDN------- FISCLVR --- LDAKFRAF -KSLD- KDGTG- QIQYK- IQEVL- QLT   CANS_BOVIN    - NYS
CANS_PIG      N D ---FDN------- FISCLVR --- LDAKFRAF -KSLD- KDGTG- QIQVN- IQEVL- QLT   CANS_PIG      - NYS
TRYT_CANFA    S DSC QGDSG ---G-- PLVCRVR GV VLQAGYVSVGEGCAQPNRPGIYTRY A YYLDVIHQYV  TRYT_CANFA    PKEP
TRYB_HUMAN    R DSC QGDSG ---G-- PLVCKYN GT VLQAGYVSVGEGCAQPNRPGIYTRY T YYLDVIHHYV  TRYB_HUMAN    PKKP
SNPA_STRLI    L NSGGGPG-------P-SCTNP-YPNSTERSRVNQLVAYGFQAALDKALEKASQR-----          SNPA_STRLI    - - - -
```

C

```
JC6052     ---N FVKLLRSVA I GL IVGA I L LYANPSLR-----------S--LNPLSTPQFDSTD--- 49
ELYA_BACAO --NKKPLGK I VASTAL L I SVAFSSSIASAAE---------EA--KE KYLI GFNEQEAVSE
SUBT_BACST MRSKKLWIS L LFA L TL I FTNAFSNNSYQAAG---------KSSTEK KYIY GF KQT-----
PRTK_TRIAL -------N RLSYL L S- LLP LA L GA P AVEQRSEAAPLIEARGEN YA N KYIY K F KEGSALSA
ALP_TRIHA ---N TSIR R L A LY L GA LLPA V L AA P AALHK---------KPEA V PN KFIYT L KEGAS-ID
ALP_CEPAC ---N YTLR R L A VL L G- A IPA A L AA P TTQKR-----------EV V PN KYIVT L KEGAS--N
ORYZ_ASPFL ---N QSIK R T LLL L GA VLPA V L AGP IFPHRR--------APIT I PG KYIVT F KSDYDQAA
ORYZ_ASPFU ---N LSIK R T LLL L GA VLPA VFGA P VQETRR-------AAQK I PG KYIVT F KPGTDTAT
I50494     ---N AVAAPHEGHDHDGH PA DHYHHLHHGKDE-------AHPSHGGEDACHLLSPHN-AD
```

```
JC6052     - E TPAS Y NL----------------------AV R RAAPAVVN VY NRGL N TNSHNQ L EIRTLG 80
ELYA_BACAO FVEQVEAHDE--VAI L SEEEE------YE IE LLHEFETIP V LSVELSPED V DAL EL D P A I
SUBT_BACST -NSAMSS A KK--K DV I SE K G---------GKVQKQFKYYN A A A T L DE KA V K EL L KD P S V
PRTK_TRIAL L D AANEK I S----------------G-KPDHVYKNV F S GF A A TL DE NN V RVL R AH P D V
ALP_TRIHA T D SH LA N VTDIH T R S L T KR S T------A GVE K IYN H TN N AYAG E FDE E T I EQI K SN P D V
ALP_CEPAC F D SH IS N VYSDIHKR G L S RR S T------A GIE K EFHI DT F N AY VG E FDE T T I EEI K NN P D V
ORYZ_ASPFL I D KHTA N ATDIHKR N L QRR DSSEEDLPI GIE R NFKI N K F A AYSG SF DE DT I AQI R QSD E V
ORYZ_ASPFU I E SH TL N ATDLHKR N L ERR DTTSGEPPV GIE K SYK I X D F A AYAG SF DD A T I EEI R KRG D V
I50494     FAFSLYKKLA L H PDAQG K N--------IFFSPYGISNALS N L A VGAKGS T L SQ I TSSLGY
```

FIG. 4—Continued.

common ancestor by divergent evolution. In general, the sequences belonging to the same family show more homology or are more closely related. This criterion is currently used to assign a particular sequence to a particular family, i.e., the serine protease, cysteine protease, aspartate protease, or metalloprotease family. Within a family, the extent of homology is inversely proportional to the phylogenetic distance. The proteases from distantly related organisms show less homology or

more diversity. However, this needs extensive sequence analysis of proteases, since the homology depends on many parameters or factors such as structure, function, source, and nature of the catalytic or active site. Thus, proteases are highly diverse enzymes having different active sites and metal-binding regions. The residues involved in disulfide bond formation and their positions, which vary in different proteases, can be detected by multiple alignments.

JC6052   S G V I N Q R G Y I I T N K - - - - - - - - - - - - - - - H V I N D A D Q - - - I I V A L Q - D - - - G - R Y F E A 116
ELYA_BACAO   S Y I E E D A E V T T K A Q S - - - - - - - - - - - - - V P N G I S R V Q A P A A H N R G L T - - - - - G - S G V K V
SUBT_BACST   A Y V E E D H I A H E Y A Q S - - - - - - - - - - - - - V P Y G I S Q I K A P A L H S Q G Y T - - - - - G - S K Y K V
PRTK_TRIAL   E Y I I E Q D A Y V T I N A A Q - - - - - - - - - - - - T N A P Y G L A R I S S T S P G T S T Y Y Y D E S A G - Q G S C V
ALP_TRIHA   A S V E P D Y I N H L S D I Y E D - - - K R A L T T Q S G A P Y G L G T V S H R T S G S T S Y I Y D S S A G - - A G T F A
ALP_CEPAC   L E V E E D Q I V H L F D E Q D E G E F S T A A L V T Q N G A V G L G T I S H R Q P G S T S Y I Y D D S A G - S G T Y A
ORYZ_ASPFL   A A V E E D Q V V H L F D L T - - - - - - - - - - - T Q S D A P Y G L G S I S H K G Q P S T D Y I Y D T N G G - E G T Y A
ORYZ_ASPFU   A H V E E D Q I V Y L D A L T - - - - - - - - - - - T Q K G A P Y G L G S I S H K G Q A S T D Y I Y D T S A G - A G T Y A
150494   S G L K A Q Q V N E G Y E H L - - - - - - - - - - - - - - I H H L G H S Q D T N Q L E A G A G V A I R - - E G F K Y V D Q

JC6052   L L V G S D S L T D L A V L K I N A T G G L P T I P I N - - A R R V P H I G D Y V L A I G N P - - - - Y N L G Q T I T Q 170
ELYA_BACAO   A V L D T G I S - T H P D L N I R G G A S F V P G E P S - T Q D G N G H G T H V A G T I A A L N N S I G V L G V A P N A
SUBT_BACST   A V I D S G I D S S H P D L N V R G G A S F V P S E T N P Y Q D G S S H G T H V A G T I A A L N N S I G V L G V S P S A
PRTK_TRIAL   Y V I D T G I E A S H P E F E G R A Q K V K T Y Y Y S S - R D G N G H G T H C A G T V G S R - - - - T Y G V A K K T
ALP_TRIHA   Y Y V D S G I N T S H Q Q F G G R A S L G Y N A A G G Q - H V D T L G H G T H V S G T I G G S - - - - T Y G V A K Q A
ALP_CEPAC   Y Y V D T G I L E S H N E F S G R A I T G Y N A V G G S - N A D T N G H G T H V A G T I G G R - - - - T Y G V A K N T
ORYZ_ASPFL   Y Y V D I G I N V D H E E F E G R A S L A Y H A A G G Q - H V D G V G H G T H V S G T I G G K - - - - T Y G V A K K A
ORYZ_ASPFU   Y Y V D S G I N V N H V E F E S R A S L A Y N A A G G S - H V D S I G H G T H V A G T I G G K - - - - T Y G V A K K T
150494   F L K D V Q H Y Y N S E A F S Y D F S K P E I A A E E I - - N Q F I A K K T N D K I T D N V K - - - - - O L D S D N Y K

JC6052   G I I S A T G R I G L N P T G R Q N F L Q T D A S I N H G H - - S - G - - G A L V N S L G E L N G I - - N T L S F D L S 223
ELYA_BACAO   E L Y A V K Y L G A S G - S G S V S S I A Q G L E V A G N N - - - - - - - G N H Y A N L S L G S P S - - P S A T L E Q A
SUBT_BACST   S L Y A V K Y L D S T G - S G Q Y S V I I N G I E V A I S N - - - - - - - N N D V I N N S L G G P S - - G S T A L K T V
PRTK_TRIAL   Q L F G V K Y L D D N G - S G Q Y S T I I A G N D F Y A S D K N H R N C P K G V Y A S L S L G G G - - - Y S S S V N S A
ALP_TRIHA   S L I S V K Y F A G E - - S A S T S V I L D G Y N V A V N D I V S K S R A S K S A I X N S L G G P - - - A S S T N T T A
ALP_CEPAC   N L I A V K Y F R G S - - S S S T S I I L D G F N V A V N D I I N R G R Q N K A A I S N S L G G G - - - Y S S A F N N A
ORYZ_ASPFL   N L L S V K Y F V G E - - S S S T S I I L D G F N V A A N D I V S K K R T G K A A I N N S L G G G - - - Y S K A F N D A
ORYZ_ASPFU   N L L S V K Y F Q G E - - S S S T S I I L D G F N V A V N D I V S K G R T K K A A I N N G L G G G - - - Y S T A F N N A
150494   N L I N Y N Y F R G K V D K P F E A Q L T H K A E F K V D K - - - - - D T I V Q V D N N K R T G R Y D I Y Q D P V N Q T

JC6052   N D G E T P - E G - - - I G F A I P F Q L A T - - - - - - - - K I N D K L I R - - D G R V I R G Y I G I G G R E I A P L 269
ELYA_BACAO   V N S A T S - R G - - - Y L V V A A S G N S G - - - - - - A G S I S Y P A R Y - - A K A N A V G A T D Q N H N - R A S F
SUBT_BACST   V D K A V S - S G - - - I V V V A A A A G N E G S S G S - - S S T V G Y P A K Y - - P S T I A V G A V N S S N Q - R A S F
PRTK_TRIAL   A A R L Q S - S G - - - Y N V A V A A G N N N A D - - - - A R N Y S P A S E - - P S V C T V G A S D R Y D R - R S S F
ALP_TRIHA   I N A A F N - Q G - - - V L T I V A A G N G D S L G N P Q P V S G T S P A N Y - - P N A I T V A A L D I K N R - T A S F
ALP_CEPAC   V N T A Y S - R G - - - V L S V V A A G N D N Q N - - - - - A A N Y S P A S A - - A N A I T Y G S I A S H N A - R S S F
ORYZ_ASPFL   V E N A F N - E G - - - V L S I Y A A G N E N T D - - - - - A S R T S P A S A - - P D A F T V A A I N V N N T - R A Y F
ORYZ_ASPFU   V E H A F D - E G - - - Y L S V Y A A G N E N S D - - - - - A S N T S P A S A - - P N A L T V A A I N K S N A - R A S F
150494   T V K N V P Y K G N T S N N I V L P D E G K N K D - - - Y E E S I C R H H L K N W H D K L F R S S V D L F N P - K F S I

FIG. 4—Continued.

FIG. 4. Homology alignment of the protease sequences. The protease sequences have been selected from the SWISS-PROT and PIR entries, and some have been deduced from the nucleotide sequences obtained from the EMBL database. These are aligned by using CLUSTAL W software for multiple alignment (291). (A) Acidic proteases. (B) Neutral proteases. (C) Alkaline proteases. The key to the sequences is given in Table 5. Numbering of the amino acid residues is based on the first sequence in the list. Identical (O) and conserved (●) residues are boxed; those involved in the active site are indicated by *.

CURRENT PROBLEMS AND POTENTIAL SOLUTIONS

Proteases are a complex group of enzymes which differ in their properties such as substrate specificity, active site, and catalytic mechanism. Their exquisite specificities provide a basis for their numerous physiological and commercial applications. Despite the extensive research on several aspects of proteases from ancient times, there are several gaps in our knowledge of these enzymes and there is tremendous scope for improving their properties to suit projected applications. The future lines of development would include (i) genetic approaches to generate microbial strains for hyperproduction of the enzymes, (ii) application of SDM to design proteases with unique specificity and increased resistance to heat and alkaline pH, (iii) synthesis of peptides (synzymes) to mimic proteases, (iv) production of abzymes (catalytic antibodies) with specific protease activity, and (v) understanding of the structure-function relationship of the enzymes. Although the section on protein engineering describes in detail how SDM has been used to alter the properties and functions of proteases of bacterial, fungal, and viral origins, some of the important problems faced in their desired usages and the possible solutions to overcome these hurdles are discussed below.

Enhancement of Thermostability

The industrial use of proteases in detergents or for leather processing requires that the enzyme be stable at higher temperatures. One of the common strategies to enhance the thermostability of the enzyme is to introduce disulfide bonds into the protease by SDM. Introduction of a disulfide bond into subtilisin E from *Bacillus subtilis* resulted in an increase of 4.5°C in the T_m of the mutant enzyme without causing any change in its catalytic efficiency (280). However, the properties of the mutant enzyme were found to revert to those of the wild-type enzyme. Enhanced stability of subtilisin was observed as a result of mutations of Asn109 and Asn218 to Ser. The analog containing both the mutations showed an additive effect on thermal stability. Thermostability of the alkaline protease from *Aspergillus oryzae* is important because of its extensive use in the manufacture of soy sauce. The optimal temperature of the wild-type enzyme was enhanced from 51 to 56°C by the introduction of a disulfide (Cys 169-Cys 200) bond. Another strategy for improving the stability of the protease was by replacing the polar amino acid groups by hydrophobic groups. The presence of positively charged amino acids in the N-terminal turn of an α-helix is undesirable in view of the possibility of an occurrence of the repulsive interactions with the helix dipole. Replacement of Lys by Ser or Asp resulted in an increase in the thermostability of the neutral protease from *B. subtilis* in the range of 0.3 to 1.2°C (54). Although these approaches result in an increased stability of proteases, the difference in the thermostabilities of the parent and the mutant enzymes is only marginal, and further research involving cassette mutagenesis, etc., is necessary to yield an enzyme with substantially enhanced thermostability.

TABLE 6. Proteases selected for multiple alignment[a]

SWISS-PROT/PIR entry	No. of amino acid residues	Type of protease	Residue(s) at active site
Acidic proteases			
CYS2_HORVU	373	C1/papain (cysteine)	C158 H297, N318
HRTD_CROAT	414	M12B (metallo)	E311
GPR_BACME	371	U3 (aspartic)	D89, D258
ASPP_AEDAE	395	Lysosomal (aspartic)	D96, D258
CARP_CANTR	390	Candidapepsin (aspartic)	D96, D258
CARP_SACFI	390	Saccharopepsin (aspartic)	D96, D258
PEPC_RAT	382	Gastricsin (aspartic)	D96, D258
PEP2_MACFU	378	Pepsin A (aspartic)	D96, D258
Neutral proteases			
PRCA_THEAC	233	PS	U
PRC3_YEAST	250	PS	U
PRC6_SCHPO	259	PS	U
PRC2_ORYSA	270	PS	U
PRC6_ARATH	250	PS	U
PRC6_DICDI	250	PS	U
PRC8_CAEEL	259	PS	U
PRC6_DROME	249	PS	U
PRC3_XENLA	233	PS	U
CANS_BOVIN	263	C2/calpain (cysteine)	U
CANS_PIG	266	C2/calpain (cysteine)	U
TRYT_CANFA	275	S1/trypsin (serine)	H74, D121, S191
TRYB_HUMAN	275	S1/trypsin (serine)	H74, D121, S191
SNPA_STRLI	237	M7 (metallo)	E64
Alkaline proteases			
JC6052	355	Trypsin-like protease	H91, D126, S201
EYLA_BACAO	380	S8/subtilase	D120, H150, S302
SUBT_BACST	381	S8/subtilase	D120, H150, S302
PRTK_TRIAL	384	S8/subtilase	D120, H150, S302
ALP_TRIHA	409	S8/subtilase	D120, H150, S302
ALP_CEPAC	402	S8/subtilase	D120, H150, S302
ORYZ_ASPFL	403	S8/subtilase	D120, H150, S302
ORYZ_ASPFU	403	S8/subtilase	D120, H150, S302
150494	410	Serine protease inhibitor	U

[a] Key to the entry names of acidic proteases: CYS2_HORVU, *Hordeum vulgare*; HRTD_CROAT, *Crotalus atrox*; GPR_BACME, *Bacillus megaterium*; ASPP_AEDAE, *Aedes aegyptii*; CARP_CANTR, *Candida tropicalis*; CARP_SACFI, *Saccharomycopsis fibuligera*; PEPC_RAT, *Rattus norvegicus*; PEP2_MACFU, *Macaca fuscata*. Sequences are numbered according to the *Hordeum vulgare* cysteine protease. Key to the neutral protease sequences: PRCA_THEAC, *Thermoplasma acidophilum*; PRC3_YEAST, *Saccharomycopsis fibuligera*; PRC6_SCHPO, *Schizosaccharomyces pombe*; PRC2_ORYSA, *Oryza sativa*; PRC6_ARATH, *Arabidopsis thaliana*; PRC6_DICDI, *Dictyostellium discoideum*; PRC8_CAEEL, *Caenorhabditis elegans*; PRC6_DROME, *Drosophila melanogaster*; PRC3_XENLA, *Xenopus laevis*; CANS_BOVIN, *Bos taurus*; CANS_PIG, *Sus scrofa*; TRYT_CANFA, *Canis familiaris*; TRYB_HUMAN, *Homo sapiens*; SNPA_STRLI, *Streptomyces lividans*. Sequences are numbered according to the *Thermoplasma* protease. PS, proteasome subunit; U, unknown. Key to the alkaline protease sequences: JC6052, *Escherichia coli*; ELYA_BACAO, *Bacillus amyloliquefaciens*; SUBT_BACST, *Bacillus subtilis*; PRTK_TRIAL, *Tritirachium album* Limber; ALP_TRIHA, *Tritirachium harzianum*; ALP_CEPAC, *Cephalosporium acremonium*; ORYZ_ASPFL, *Aspergillus flavus*; ORYZ_ASPFU, *Aspergillus fumigatus*; 150494, *Cyprinus carpio*. Residues are numbered according to the *E. coli* protease.

Prevention of Autoproteolytic Inactivation

Subtilisin, an extensively studied protease, is widely used in detergent formulations due to its stability at alkaline pH. However, its autolytic digestion presents a major problem for its use in industry. It was deduced that there is a correlation between the autolytic and conformational stabilities. Computer modelling followed by introduction of a Cys24 or Cys87 mutation resulted in destabilization of subtilisin from *Bacillus amyloliquefaciens* (312). Introduction of a disulfide bond increased the stability of the mutant to a level less than or equal to that of the wild-type enzyme. It appears logical that mutations in the amino acids involved at the site of autoproteolysis may prevent the protease inactivation caused by self-digestion.

Alteration of pH Optimum

Different applications of proteases require specific optimal pHs for the best performance of the enzyme; e.g., the use of proteases in the leather and detergent industries requires an enzyme with an alkaline pH optimum, whereas the use in the cheesemaking industry requires an acidic protease. Protein engineering enables us to tailor the pH dependence of the enzyme catalysis to optimize the industrial processes. Modifications in the overall surface charge of the proteins are known to alter the optimal pH of the enzyme. A change of Asp99 to Ser in subtilisin from *Bacillus amyloliquefaciens* has demonstrated the potential of altering the optimal pH of the enzyme by systematic multiple mutations on the surface of the protein (290).

Changing of Substrate Specificity

The properties needed for industrial applications of proteases differ from their physiological properties. The natural substrates of the enzyme are usually different from those desired for their industrial applications. Despite extensive research on proteases, relatively little is known about the factors that control their specificities toward nonphysiological sub-
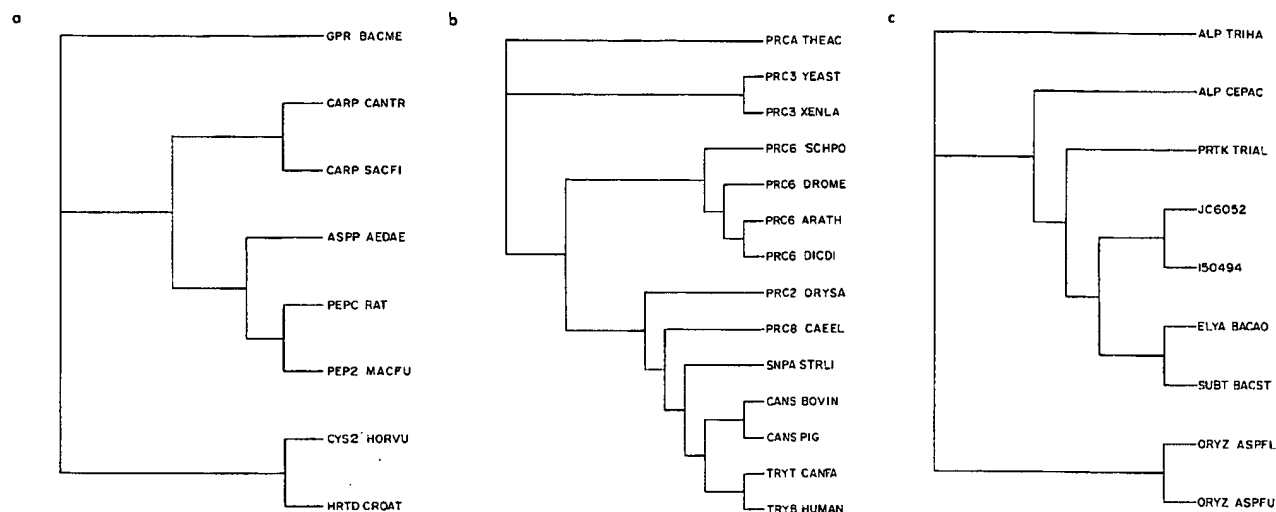
FIG. 5. Dendrogram showing the relationships among the proteases, created by the TreeView package (213). The proteases are grouped as acidic proteases (a), neutral proteases (b), and alkaline proteases (c). Abbreviations of the species described are those used in Table 6. The differences between the sequences are proportional to the length along the horizontal axis.

strates. Strategies involving SDM are being explored to tailor these specificities at will. A combinatorial random-mutagenesis approach has been used to generate mutants that secrete proteases with functional properties different from those of the parent enzyme (77). Introduction of several point mutations into the substrate-binding site of α-lytic protease was shown to affect its specificity in a predictable manner. The protease preferentially cleaves on the C-terminal side of small uncharged residues such as Ala, mainly because the pocket that accommodates the substrate P-1 residue is shallow due to the presence of two bulky methionine residues (Met190 and Met213) at the subsite. Replacement of Met213 with a His residue had a beneficial effect on its substrate specificity.

Improvement of Yield

The cost of enzyme production is a major obstacle in the successful application of proteases in industry. Protease yields have been improved by screening for hyperproducing strains and/or by optimization of the fermentation medium. Strain improvement by either conventional mutagenesis or recombinant-DNA technology have been useful in improving the production of proteases. Hyperexpression by genetic manipulation of microbes is described in the section on genetic engineering. Increases in the yield of viral proteases are particularly important for developing therapeutic agents against devastating diseases such as malaria, cancer, and AIDS.

There are many major problems in the commercialization of proteases. Although they are being addressed by both conventional and novel methods of genetic manipulation, there are no entirely satisfactory solutions and many of these problems remain unanswered.

FUTURE SCOPE

Proteases are a unique class of enzymes, since they are of immense physiological as well as commercial importance. They possess both degradative and synthetic properties. Since proteases are physiologically necessary, they occur ubiquitously in animals, plants, and microbes. However, microbes are a gold-mine of proteases and represent the preferred source of en-

zymes in view of their rapid growth, limited space required for cultivation, and ready accessibility to genetic manipulation. Microbial proteases have been extensively used in the food, dairy and detergent industries since ancient times. There is a renewed interest in proteases as targets for developing therapeutic agents against relentlessly spreading fatal diseases such as cancer, malaria, and AIDS. Advances in genetic manipulation of microorganisms by SDM of the cloned gene opens new possibilities for the introduction of predesigned changes, resulting in the production of tailor-made proteases with novel and desirable properties. The development of recombinant rennin and its commercialization by Pfizer and Genencor is an excellent example of the successful application of modern biology to biotechnology. The advent of techniques for rapid sequencing of cloned DNA has yielded an explosive increase in protease sequence information. Analysis of sequences for acidic, alkaline, and neutral proteases has provided new insights into the evolutionary relationships of proteases.

Despite the systematic application of recombinant technology and protein engineering to alter the properties of proteases, it has not been possible to obtain microbial proteases that are ideal for their biotechnological applications. Industrial applications of proteases have posed several problems and challenges for their further improvements. The biodiversity represents an invaluable resource for biotechnological innovations and plays an important role in the search for improved strains of microorganisms used in the industry. A recent trend has involved conducting industrial reactions with enzymes reaped from exotic microorganisms that inhabit hot waters, freezing Arctic waters, saline waters, or extremely acidic or alkaline habitats. The proteases isolated from extremophilic organisms are likely to mimic some of the unnatural properties of the enzymes that are desirable for their commercial applications. Exploitation of biodiversity to provide microorganisms that produce proteases well suited for their diverse applications is considered to be one of the most promising future alternatives. Introduction of extremophilic proteases into industrial processes is hampered by the difficulties encountered in growing the extremophiles as laboratory cultures. Revolutionary robotic approaches such as DNA shuffling are being

developed to rationalize the use of enzymes from extremophiles. The existing knowledge about the structure-function relationship of proteases, coupled with gene-shuffling techniques, promises a fair chance of success, in the near future, in evolving proteases that were never made in nature and that would meet the requirements of the multitude of protease applications.

A century after the pioneering work of Louis Pasteur, the science of microbiology has reached its pinnacle. In a relatively short time, modern biotechnology has grown dramatically from a laboratory curiosity to a commercial activity. Advances in microbiology and biotechnology have created a favorable niche for the development of proteases and will continue to facilitate their applications to provide a sustainable environment for mankind and to improve the quality of human life.

ACKNOWLEDGMENTS

REFERENCES

1. **Akimkina, T. V., E. A. Nosovskaya, and S. V. Kostrov.** 1992. Cloning and expression of the neutral proteinase gene of *Bacillus cereus* in *Bacillus subtilis* cells. Mol. Biol. **26:**418–423.
2. **Allen, C., V. K. Stromer, F. D. Smith, G. H. Lacy, and M. S. Mount.** 1986. Complementation of an *Erwinia carotovora* subsp. *carotovora* protease mutant with a protease encoding cosmid. Mol. Gen. Genet. **202:**276–279.
3. **Amerik, A., L. G. Chistyakova, N. I. Ostroumova, A. I. Gurevich, and V. K. Antonov.** 1988. Cloning, expression and structure of the truncated functionally active *lon* gene of *Escherichia coli*. Bioorg. Khim. **14:**408–411.
4. **Ammerer, G., C. P. Hunter, J. H. Rothman, G. C. Saari, L. A. Valls, and T. M. Stevens.** 1986. PEP4 gene of *Saccharomyces cerevisiae* encodes proteinase A, a vacuolar enzyme required for processing of vacuolar precursors. Mol. Cell. Biol. **6:**2490–2497.
5. **Argos, P.** 1987. A sensitive procedure to compare amino acid sequences. J. Mol. Biol. **193:**385–396.
6. **Argos, P., G. Kamer, M. J. H. Nicklin, and E. Wimmer.** 1984. Similarity in gene organization and homology between proteins of animal picornaviruses and a plant comovirus suggest common ancestry of these virus families. Nucleic Acids Res. **12:**7251–7267.
7. **Atsumi, Y., S. Yamamoto, K. Morihara, J. Fukushima, H. Takeuchi, N. Mizuki, S. Kawamoto, and K. Okuda.** 1989. Cloning and expression of the alkaline proteinase gene from *Pseudomonas aeruginosa* IFO 3455. J. Bacteriol. **171:**5173–5175.
8. **Austew, B. M., and E. L. Smith.** 1976. Action of staphylococcal proteinase on peptides of varying chain length and composition. Biochem. Biophys. Res. Commun. **72:**411–417.
9. **Avakov, A. S., A. P. Bolotin, and S. V. Sorokin.** 1990. The structure of *Bacillus brevis* metalloprotease gene. Mol. Biol. **24:**1363–1372.
10. **Ballinger, M. D., J. Tom, and J. A. Wells.** 1996. Furilisin: a variant of subtilisin BPN' engineered for cleaving tribasic substrates. Biochemistry **35:**13579–13585.
11. **Baker, E. N., and J. Drenth.** 1987. Active sites of enzymes, p. 314. *In* F. A. Jurnak and A. McPherson (ed.), Biological macromolecules and assemblies, vol. 3. Active sites of enzymes. John Wiley & Sons, Inc., New York, N.Y.
12. **Barett, A. J.** 1994. Proteolytic enzymes: serine and cysteine peptidases. Methods Enzymol. **244:**1–15.
13. **Barett, A. J.** 1995. Proteolytic enzymes: aspartic and metallopeptidases. Methods Enzymol. **248:**183.
14. **Berglund, P., M. R. Stabile, M. Gold, J. B. Jones, C. Mitchinson, R. R. Bolt, and T. P. Graycar.** 1996. Altering the specificity of subtilisin *B. lentus* by combining site-directed mutagenesis and chemical modification. Bioorg. Med. Chem. Lett. **6:**2507–2512.
15. **Berka, R. M., M. Ward, L. J. Wilson, K. J. Hayenga, K. H. Kodama, L. P. Carlomagno, and S. A. Thompson.** 1990. Molecular cloning and deletion of the gene encoding aspergillopepsin A from *Aspergillus awamori*. Gene **86:**153–162.
16. **Bhosale, S. H., M. B. Rao, V. V. Deshpande, and M. C. Srinivasan.** 1995. Thermostability of high activity alkaline protease from *Conidiobolus coro-*

nanus (NCL 86.8.20). Enzyme Microb. Technol. **17:**136–139.
17. **Binnie, C., L. Liao, E. Walczyk, and L. T. Malek.** 1996. Isolation and characterization of a gene encoding a chymotrypsin-like serine protease from *Streptomyces lividans* 66. Can. J. Microbiol. **42:**284–288.
18. **Binnie, C., M. J. Butler, J. S. Aphale, R. Bourgault, M. A. Dizonno, P. Krygsman, L. Liao, E. Walczyk, and L. T. Malek.** 1995. Isolation and characterization of 2 genes encoding proteases associated with the mycelium of *Streptomyces lividans* 66. J. Bacteriol. **177:**6033–6040.
19. **Blobel, G., and B. Dobberstein.** 1975. Transfer of proteins across membranes. J. Cell Biol. **67:**835–851.
20. **Blundell, T. L., J. B. Cooper, A. Sali, and Z. Zhu.** 1991. Comparisons of the sequences, 3-D structures and mechanisms of pepsin-like and retroviral aspartic proteinases. Adv. Exp. Med. Biol. **306:**443–453.
21. **Boguslawski, G., J. L. Shultz, and C. O. Yehle.** 1983. Purification and characterization of an extracellular protease from *Flavobacterium arborescens*. Anal. Biochem. **132:**41–49.
22. **Boye, K., B. M. Stummann, and K. W. Henningsen.** 1992. cDNA cloning and sequencing of the bean yellow mosaic virus nuclear inclusion protein genes. Plant Mol. Biol. **18:**1203–1205.
23. **Boyer, P. D.** 1971. The enzymes, 3rd ed. Academic Press, Inc., New York, N.Y.
24. **Braunagel, S. C., and M. J. Benedik.** 1990. The metalloprotease gene of *Serratia marcescens* strain SM6. Mol. Gen. Genet. **222:**446–451.
25. **Brenner, S.** 1988. The molecular evolution of genes and proteins: a tale of two serines. Nature **334:**528–530.
26. **Browner, M. F., W. W. Smith, and A. L. Castelhano.** 1995. Matrilysin-inhibitor complexes: common themes among metalloproteinases. Biochemistry **34:**6601–6610.
27. **Brueckner, R., O. Shoseyor, and R. H. Doi.** 1990. Multiple active forms of a novel serine protease from *Bacillus subtilisin*. Mol. Gen. Genet. **221:**486–490.
28. **Bruinenberg, P. G., W. M. De Vos, and R. J. Siezen.** 1994. Prevention of C-terminal autoprocessing of *Lactococcus lactis* SK11 cell-envelope proteinase by engineering of an essential surface loop. Biochem. J. **302:**957–963.
29. **Carmona, C., and G. L. Gray.** 1987. Nucleotide sequence of the serine protease gene of *Staphylococcus aureus*, strain V8. Nucleic Acids Res. **15:**6757.
30. **Carter, P., and J. A. Wells.** 1987. Engineering substrate specificity by substrate assisted catalysis. Science **237:**394–399.
31. **Cerny, G.** 1978. Studies on the aminopeptidase test for the distinction of gram-negative from gram-positive bacteria. Eur. J. Appl. Microbiol. Biotechnol. **5:**113–122.
32. **Chang, P. C., T.-C. Kuo, A. Tsugita, and Y.-H. W. Lee.** 1990. Extracellular metalloprotease gene of *Streptomyces cacaoi*: structure, nucleotide sequence and characterization of the cloned gene product. Gene **85:**87–95.
33. **Cheevadhanarak, S., D. V. Renno, G. Saunders, and G. Holt.** 1991. Cloning and selective overexpression of an alkaline protease-encoding gene from *Aspergillus oryzae*. Gene **108:**151–155.
34. **Cheng, J.-C., C.-P. Shao, and L.-I. Hor.** 1996. Cloning and nucleotide sequencing of the protease gene of *Vibrio vulnificus*. Gene **183:**255–257.
35. **Chin, D. T., S. A. Goff, T. Webster, T. Smith, and A. L. Goldberg.** 1988. Sequence of the *lon* gene in *Escherichia coli*. A heat-shock gene which encodes the ATP-dependent protease La. J. Biol. Chem. **263:**11718–11728.
36. **Chiplonkar, J. M., S. V. Gangodkar, U. V. Wagh, G. D. Ghadge, M. V. Rele, and M. C. Srinivasan.** 1985. Applications of alkaline protease from *Conidiobolus* in animal cell culture. Biotechnol. Lett. **7:**665–668.
37. **Cho, W.-L., and A. S. Raikhel.** 1992. Cloning of cDNA for mosquito lysosomal aspartic protease. J. Biol. Chem. **267:**21823–21829.
38. **Chung, C. H., and A. L. Goldberg.** 1981. The product of the *lon* (*capR*) gene in *Escherichia coli* is the ATP dependent protease, protease La. Proc. Natl. Acad. Sci. USA **78:**4931–4935.
39. **Cooper, J. B., G. Khan, G. Taylor, I. J. Yickle, and T. L. Blundell.** 1990. X-ray analyses of aspartic proteinases. II. Three dimensional structure of the hexagonal crystal form of porcine pepsin at 2.3 A resolution. J. Mol. Biol. **214:**199–222.
40. **Corr, B. R., M. Richardson, and A. F. Wilderspin.** 1992. Cloning, expression and mutagenesis of SIVmac proteinase in *E. coli*. Biochem. Soc. Trans. **20:**1605.
41. **Dancer, B. N., and J. Mandelstam.** 1975. Production and possible function of serine protease during sporulation of *Bacillus subtilis*. J. Bacteriol. **121:**406–410.
42. **Date, T., and W. Wickner.** 1981. Isolation of the *Escherichia coli* leader peptidase gene and effects of leader peptidase overproduction *in vitro*. Proc. Natl. Acad. Sci. USA **78:**6106–6110.
43. **David, V. A., A. H. Deutch, A. Sloma, D. Pawlyk, A. Ally, and D. R. Durham.** 1992. Cloning, sequencing and expression of the gene encoding the extracellular neutral protease, vibriolysin, of *Vibrio proteolyticus*. Gene **112:**107–112.
44. **Davidow, L. S., M. M. O'Donnell, F. S. Kaczmaerk, D. A. Pereira, J. R. DeZeeuw, and A. E. Franke.** 1987. Cloning and sequencing of the alkaline

extracellular protease gene of *Yarrowia lipolytica*. J. Bacteriol. **169:**4621–4629.

45. **Deane, S. M., F. T. Robb, S. M. Robb, and D. R. Wood.** 1989. Nucleotide sequence of the *Vibrio anginolyticus* calcium-dependent, detergent-resistant alkaline serine exoprotease A. Gene **76:**281–288.

46. **Debouck, C., J. G. Gorniak, J. E. Strickler, T. D. Meek, B. W. Metcalf, and M. Rosenberg.** 1987. Human immunodeficiency virus protease expressed in *Escherichia coli* exhibits autoprocessing and specific maturation of the gag precursor. Proc. Natl. Acad. Sci. USA **84:**8903–8906.

47. **Deckman, I. C., M. Hagen, and P. J. McCann.** 1992. Herpes simplex virus type I protease expressed in *Escherichia coli* exhibits autoprocessing and specific cleavage of the ICP 35 assembly protein. J. Virol. **66:**7362–7367.

48. **De Marco, A. C., and A. J. Dick.** 1978. Aminopeptidase I activities in several microorganisms. Can. J. Biochem. **56:**66–71.

49. **De Vos, W. M.** 1987. Gene cloning and expression in lactic streptococci. FEMS Microbiol. Rev. **46:**281–295.

50. **De Vos, W. M., P. Vos, H. deHaard, and I. Boerrigter.** 1989. Cloning and expression of the *Lactococcus lactis* subsp. *cremoris* SkI I gene encoding an extracellular serine proteinase. Gene **85:**169–176.

51. **Dickinson, L., M. Harboe, R. Van Heeswijck, P. Stroeman, and L. P. Jepsen.** 1987. Expression of active *Mucor miehei* aspartic protease in *Mucor circinelloides*. Carlsberg Res. Commun. **52:**243–252.

52. **Drapeau, G. R., Y. Boily, and J. Houmard.** 1972. Purification and properties of an extracellular protease of *Staphylococcus aureus*. J. Biol. Chem. **247:**6720–6726.

53. **Egel-Mitani, M., H. P. Flygering, and M. T. Hansen.** 1990. A novel aspartyl protease allowing KEX-2 independent MFα propheromone processing in yeast. Yeast **6:**127–137.

54. **Eijsink, V. G. H., G. Vriend, B. vanden Burg, J. R. vander Zee, and G. Venema.** 1992. Increasing the thermostability of a neutral protease by replacing positively charged amino acids in the N-terminal turn of α-helices. Protein Eng. **5:**165–170.

55. **Eijsink, V. G. H., J. K. vander Zee, B. vanden Burg, G. Vriend, and G. Venema.** 1991. Improving the thermostability of the neutral protease of *Bacillus stearothermophilus* by replacing a buried aspargine by leucine. FEBS Lett. **282:**13–16.

56. **Reference deleted.**

57. **Epstein, D. M., and P. C. Wensink.** 1988. The α-lytic protease gene of *Lysobacter enzymogenes*. J. Biol. Chem. **263:**16586–16590.

58. **Esposito, R. E., and S. Klapholz.** 1981. Meiosis and ascospore development, p. 211–287. *In* J. N. Strathern, E. W. Jones, and J. R. Broach, (ed.), The molecular biology of the yeast *Saccharomyces*. Life cycle and inheritance. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

59. **Estell, D. A., T. P. Gyaycar, J. V. Miller, D. B. Powers, J. P. Burnier, P. G. Ng, and J. A. Wells.** 1986. Probing steric and hydrophobic effects on enzyme substrate interactions by protein engineering. Science **233:**659–663.

60. **Fastrez, J., and A. R. Fersht.** 1973. Demonstration of the acyl-enzyme mechanism for the hydrolysis of peptides and anilides by chymotrypsin. Biochemistry **12:**2025–2034.

61. **Felix, F., and N. Brouillet.** 1966. Purification and properties of two peptidases from brewer's yeast. Biochim. Biophys. Acta **122:**127–144.

62. **Fishman, Y., J. Bricker, J. V. Gilbert, A. G. Plant, and A. Wright.** 1985. Cloning of the type 1 immunoglobulin A1 protease from *Neisseria gonorrhoeae* and secretion of the enzyme from *Escherichia coli*, p. 164–168. *In* G. K. Schoolnik (ed.), Pathogenic *Neisseria*. Proceedings of the Fourth International Symposium. American Society for Microbiology, Washington, D.C.

63. **Fitzgerald, P. M. D., B. M. McKeever, J. F. Van Middlesworth, J. P. Springer, J. C. Heimbach, Chih-Tai Leu, W. K. Herber, R. A. F. Dixon, and P. L. Darke.** 1990. Crystallographic analysis of a complex between human immunodeficiency virus type 1 protease and acetyl-pepstatin at 2.0A resolution. J. Biol. Chem. **265:**14209–14219.

64. **Fox, J. W., J. D. Shannon, and J. B. Bjarnason.** 1991. Proteinases and their inhibitors in biotechnology. Enzymes in biomass conversion. ACS Symp. Ser. **460:**62–79.

65. **Friedman, H., M. Goebel, and M. Synder.** 1992. A homolog of the proteasome-related RING10 gene is essential for yeast cell growth. Gene **122:**203–206.

66. **Fujii, G., K. Tashiro, Y. Emori, K. Saigo, K. Tanaka, and K. Shiokawa.** 1991. Deduced primary structure of a *Xenopus* proteasome subunit XC3 and expression of its mRNA during early development. Biochem. Biophys. Res. Commun. **178:**1233–1239.

67. **Fujii, M., M. Takagi, T. Imanaka, and S. Aiba.** 1983. Molecular cloning of a thermostable neutral protease gene from *Bacillus stearothermophilus* in a vector plasmid and its expression in *Bacillus stearothermophilus* and *Bacillus subtilis*. J. Bacteriol. **154:**831–837.

68. **Geisen, R.** 1993. Cloning of a protease gene from *Pseudomonas nalgiovense* by expression in *Escherichia coli*. Lett. Appl. Microbiol. **16:**303–306.

69. **Gilbert, C., D. Atlan, B. Blanc, R. Portalier, J. E. Germond, L. Lapierre, and B. Mollet.** 1996. A new cell surface proteinase: sequencing and analysis of the *prtB* gene from *Lactobacillus delbrueckii* subsp. *bulgaricus*. J. Bacteriol. **178:**3059–3065.

70. **Gilbert, J. V., A. G. Plant, Y. Fishman, and A. Wright.** 1988. Cloning of the gene encoding streptococcal immunoglobulin A1 protease and its expression in *Escherichia coli*. Infect. Immun. **56:**1961–1966.

71. **Gilles, A. M., J. M. Imhoff, and B. Keil.** 1979. α-Clostripain. Chemical characterization, activity and thiol content of the highly active form of clostripain. J. Biol. Chem. **254:**1462–1468.

72. **Godfrey, T., and S. West.** 1996. Industrial enzymology, 2nd ed., p. 3. Macmillan Publishers Inc., New York, N.Y.

73. **Goldberg, A. L., K. H. Shreedharaswamy, C. H. Chung, and F. S. Larimore.** 1983. Proteases of *Escherichia coli*. Methods Enzymol. **80:**680–702.

74. **Gomi, K., K. Arikawa, N. Kamiya, K. Kitamoto, and C. Kumagai.** 1993. Cloning and nucleotide sequence of the acid protease encoding gene (*pep A*) from *Aspergillus oryzae*. Biosci. Biotechnol. Biochem. **57:**1095–1100.

75. **Gottesman, S., and M. R. Maurizi.** 1992. Regulation by proteolysis: energy-dependent proteases and their targets. Microbiol. Rev. **56:**592–621.

76. **Govind, N. S., B. Mehta, M. Sharma, and V. V. Modi.** 1981. Protease and carotenogenesis in *Blakeslea trispora*. Phytochemistry **20:**2483–2485.

77. **Graham, L. D., K. D. Haggett, P. A. Jennings, D. S. Brocque, and R. G. Whittaker.** 1993. Random mutagenesis of the substrate binding site of a serine protease can generate enzymes with increased activities and altered primary specificities. Biochemistry **32:**6250–6258.

78. **Graves, M. C., J. J. Lim, E. P. Heimer, and R. A. Kramer.** 1988. An 11-kDa form of human immunodeficiency virus protease expressed in *Escherichia coli* is sufficient for enzymatic activity. Proc. Natl. Acad. Sci. USA **85:**2449–2453.

79. **Gray, G. L., K. Hayenga, D. Cullen, L. J. Wilson, and S. Norton.** 1986. Primary structure of *Mucor miehei* aspartyl protease: evidence for a zymogen intermediate. Gene **48:**41–53.

80. **Grodberg, J., M. D. Lundrigan, D. L. Toledo, W. F. Mangel, and J. J. Dunn.** 1988. Complete nucleotide sequence and deduced amino acid sequence of the *omp* T gene of *Escherichia coli* K-12. Nucleic Acids Res. **16:**1209.

81. **Gunkel, P. A., and H. G. Gassen.** 1989. Proteinase K from *Tritirachium album Limber*: characterization of the chromosomal gene and expression of the cDNA in *Escherichia coli*. Eur. J. Biochem. **179:**185–194.

82. **Gustin, J. K., E. Kessler, and D. E. Ohman.** 1996. A substitution at His-120 in the LasA protease of *Pseudomonas aeruginosa* blocks enzyme activity without affecting propeptide processing or extracellular secretion. J. Bacteriol. **178:**6608–6617.

83. **Guzzo, J., M. Murgier, A. Filloux, and A. Lazdunski.** 1990. Cloning of the *Pseudomonas aeruginosa* alkaline protease gene and secretion of the protease into the medium by *Escherichia coli*. J. Bacteriol. **172:**942–948.

84. **Haass, C., B. Pesold-Hurt, G. Multhaup, K. Beyreuther, and P.-M. Kloetzel.** 1990. The *Drosophila* PROS-28.1 gene is a member of the proteasome gene family. Gene **90:**235–241.

85. **Hartley, B. S.** 1960. Proteolytic enzymes. Annu. Rev. Biochem. **29:**45–72.

86. **Hase, C. C., and R. A. Finkelstein.** 1991. Cloning and nucleotide sequence of the *Vibrio cholerae* hemagglutinin/protease (HA/protease) gene and construction of an HA/protease negative strain. J. Bacteriol. **173:**3311–3317.

87. **Hatanaka, M., and S. H. Nam.** 1989. Identification of HTLV-I gag protease fits sequential processing of the gag gene product. J. Cell. Biochem. 1989. **40:**15–30.

88. **He, C., M. Xu, J. Jin, and D. Reu.** 1988. Cloning and expression of an alkaline protease gene from *Bacillus licheniformis* in *Bacillus subtilis*. Gongye Weishengwu **18:**1–5.

89. **Henderson, G., P. Krygsman, C. J. Liu, C. C. Davey, and L. T. Malek.** 1987. Characterization and structure of genes for proteases A and B from *Streptomyces griseus*. J. Bacteriol. **169:**3778–3784.

90. **Henner, D. J., M. Yang, L. Band, and J. A. Wells.** 1985. Expression of cloned protease genes in *Bacillus subtilis*, p. 95–103. *In* Proceedings of the 9th International Spore Conference.

91. **Hershko, A., E. Leshinsky, D. Ganoth, and H. Heller.** 1984. ATP-dependent degradation of ubiquitin protein conjugates. Proc. Natl. Acad. Sci. USA **81:**1619–1623.

92. **Hibbs, M. S., K. A. Hasty, J. M. Seyer, A. H. Kang, and C. L. Mainardi.** 1985. Biochemical and immunological characterization of the secreted forms of human neutrophil gelatinase. J. Biol. Chem. **260:**2493–2500.

93. **Hilt, W., and D. H. Wolf.** 1996. Proteasomes: destruction as a programme. Trends Biochem. Sci. **21:**81–119.

94. **Hiramatsu, R., J. Aikawa, S. Horinouchi, and T. Beppu.** 1989. Secretion by yeast of the zymogen form of *Mucor* rennin, an aspartic proteinase of *Mucor pusillus*, and its conversion to the mature form. J. Biol. Chem. **264:**16862–16866.

95. **Hirata, D., S. Fukui, and I. Yamashita.** 1988. Nucleotide sequences of the secretable acid protease gene PEP1 in the yeast *Saccharomycopsis fubiligera*. Agric. Biol. Chem. **52:**2647–2649.

96. **Hodgson, J.** 1994. The changing bulk catalysis market: recombinant DNA techniques have changed bulk enzyme production dramatically. Bio/Technology **12:**789–790.

97. **Hoffman, T.** 1974. Food related enzymes. Adv. Chem. Ser. **136:**146–185.

98. **Holmes, M. A., and B. W. Matthews.** 1981. Binding of hydroxamic acid inhibitors to crystalline thermolysine suggests a pentacoordinate zinc intermediate in catalysis. Biochemistry **20:**6912–6920.

99. Honjo, M., K. Manabe, H. Shimada, I. Mita, A. Nakayama, and Y. Furutani. 1984. Cloning and expression of the gene for neutral protease of *Bacillus amyloliquefaciens* in *Bacillus subtilis*. J. Biotechnol. **1**:265–77.

100. Horiuchi, H., K. Yanai, T. Okazaki, M. Takagi, and K. Yano. 1988. Isolation and sequencing of a genomic clone encoding the aspartic proteinase of *Rhizopus niveus*. J. Bacteriol. **170**:272–278.

101. Horiuchi, H., T. Ashikari, T. Amachi, H. Yoshizumi, M. Takagi, and K. Yano. 1990. High level secretion of a *Rhizopus niveus* aspartic proteinase in *Saccharomyces cerevisiae*. Agric. Biol. Chem. **54**:1771–1779.

102. Houde, A., and J. M. Weber. 1987. Sequence of the protease of human subgroup E adenovirus type 4. Gene **54**:51–56.

103. Houde, A., and J. M. Weber. 1988. The primary structure of human adenovirus type 12 protease. Nucleic Acids Res. **16**:7195.

104. Houde, A., and J. M. Weber. 1988. Sequence of the human adenovirus type 3 protease. Nucleic Acids Res. **16**:11374.

105. Hruskova-Heidingsfeldova, O., M. Andreansky, M. Fabry, I. Blaha, P. Strop, and E. Hunter. 1995. Cloning, bacterial expression, and characterization of the Mason-Pfizer monkey virus proteinase. J. Biol. Chem. **270**:15053–15058.

106. Hube, B., C. J. Turver, F. C. Odds, H. Eiffert, G. J. Boulnois, H. Kochel, and R. Ruchel. 1991. Sequence of the *Candida albicans* gene encoding the secretory aspartate protease. J. Med. Vet. Mycol. **29**:129.

107. Hwang, D. H., J. C. Kim, J. H. Ko, M. J. Chun, and S. M. Byun. 1993. Molecular cloning and nucleotide sequence of the protease B gene from *Streptomyces griseus* ATCC 10137. Korean Biochem. J. **26**:383–390.

108. Ichihara, S., T. Suzuki, M. Suzuki, and S. Mizushima. 1989. Molecular cloning and sequencing of the *sppA* gene and characterization of the encoded protease IV, a signal peptide peptidase, of *Escherichia coli*. J. Biol. Chem. **261**:9405–9411.

109. Igarashi, H., T. Morita, and S. Iwanaga. 1979. A new method for purification of staphylocoagulase by a bovine prothrombin-Sepharose column. J. Biochem. **86**:1615–1618.

110. Ikegaya, K., Y. Ishida, K. Murakami, A. Masaki, N. Sugio, K. Takechi, S. Murakami, H. Tatsumi, Y. Ogawa, E. Nakano, H. Motai, and H. Kawabe. 1992. Enhancement of the thermostability of the alkaline protease from *Aspergillus oryzae* by introduction of a disulfide bond. Biosci. Biotechnol. Biochem. **56**:326–327.

111. Ikemura, H., H. Takagi, and M. Inouye. 1987. Requirement of pro-sequence for the production of active subtilisin E in *Escherichia coli*. J. Biol. Chem. **242**:7859–7864.

112. Imanaka, T., M. Shibazaki, and M. Takagi. 1986. A new way of enhancing the thermostability of proteases. Nature **324**:695–697.

113. Inoue, H., J. Lu, S. B. P. Athanda, K. Kong, T. Hayashi, T. Kimura, O. Makabe, and K. Takahashi. 1995. *Aspergillus niger* var. *macrosporus* proteinase B. cDNA cloning, expression and activation of the proenzyme. Adv. Exp. Med. Biol. **362**:581–587.

114. Inoue, H., T. Kimura, O. Makabe, and K. Takahashi. 1991. The gene and deduced protein sequences of the zymogen of *Aspergillus niger* acid proteinase A. J. Biol. Chem. **266**:19484–19489.

114a.International Union of Biochemistry. 1992. Enzyme nomenclature. Academic Press, Inc., Orlando, Fla.

115. Isogai, T., M. Fukagawa, H. Kojo, M. Kohsaka, H. Aoki, and H. Imanaka. 1991. Cloning and nucleotide sequences of the complementary and genomic DNAs for the alkaline protease from *Acremonium chrysogenum*. Agric. Biol. Chem. **55**:471–477.

116. Ito, K., S.-I. Sakakibara, and K. Izaki. 1989. Cloning of two protease genes from *Rhodocyclus gelatinosa* APR 3-2 and their expression in *Escherichia coli*. Curr. Microbiol. **18**:41–45.

117. Ivanoff, L. A., T. Towatari, J. Ray, B. D. Korant, and S. R. Patteway, Jr. 1986. Expression and site specific mutagenesis of the poliovirus 3C protease in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **83**:5392–5396.

118. Jackson, D. P., and D. A. Cotter. 1984. Expression of proteolytic enzymes during *Dictyostelium discoideum* spore germination. Arch. Microbiol. **137**:205–208.

119. Jacobs, M., M. Eliasson, M. Uhlen, and J.-I. Flock. 1985. Cloning, sequencing and expression of subtilisin Carlsberg from *Bacillus licheniformis*. Nucleic Acids Res. **13**:8913–8926.

120. James, M. N. G., A. R. Sielecki, K. Hayakawa, and M. H. Gelb. 1992. Crystallographic analysis of transition state mimics bound to penicillopepsin: difluorostatine- and difluorostatone-containing peptides. Biochemistry **31**:3872–3886.

121. James, M. N. G., I. N. Hsu, and L. T. J. Delbaere. 1977. Mechanism of acid protease catalysis based on the crystal structure of penicillopepsin. Nature **267**:808–813.

122. Jany, K. D., G. Lederer, and B. Mayer. 1986. Amino acid sequence of proteinase K from the mold *Tritirachium album* Limber. FEBS Lett. **199**:139–144.

123. Jaton-Ogay, K., M. Suter, R. Crameri, R. Falchetto, A. Faitu, and M. Monod. 1992. Nucleotide sequence of a genomic and cDNA clone encoding an extracellular alkaline protease of *Aspergillus fumigatus*. FEMS Microbiol. Lett. **92**:162–168.

124. Jaton-Ogay, K., S. Paris, M. Huerre, M. Quadroni, R. Falchetto, G. Togri,

J. Latge, and M. Monod. 1994. Cloning and disruption of the gene encoding an extracellular metalloprotease of *Aspergillus fumigatus*. Mol. Microbiol. **14**:917–928.

125. Kagani, N., H. Inone, T. Kimura, O. Makabe, and K. Takahashi. 1995. Expression in *Escherichia coli* of *Aspergillus niger* var. *macrosporus* proteinase A, a non-pepsin type acid proteinase. Adv. Exp. Med. Biol. **362**:597–603.

126. Kageyama, T., K. Tanabe, and O. Koiwai. 1991. Development dependent expression of isozymogens of monkey pepsinogens and structural differences between them. Eur. J. Biochem. **202**:205–215.

127. Kakudo, S., N. Kikuchi, K. Kitadokoro, T. Fujiwara, E. Nakamura, H. Okamoto, M. Shin, M. Tamaki, H. Teraoka, H. Tsuzuki, and N. Yoshida. 1992. Purification, characterization, cloning and expression of a glutamic acid-specific protease from *Bacillus licheniformis* ATCC 14580. J. Biol. Chem. **33**:23782–23788.

128. Kalisz, M. H. 1988. Microbial proteinases. Adv. Biochem. Eng. Biotechnol. **36**:17–55.

129. Kaneko, R., N. Koyama, Y.-C. Tsai, R.-Y. Juang, K. Yoda, and M. Yamasaki. 1989. Molecular cloning of the structural gene for alkaline elastase YaB, a new subtilisin produced by an alkalophilic *Bacillus* strain. J. Bacteriol. **171**:5232–5236.

130. Katoh, I., Y. Yashinaka, A. Rein, M. Shibuya, T. Okada, and S. Oroszlan. 1985. Murine leukemia virus maturation: protease region required for conversion from 'immature' to 'mature' core form and for virus infectivity. Virology **145**:280–292.

131. Katz, M. E., R. N. Rice, and B. F. Cheetham. 1994. Isolation and characterization of an *Aspergillus nidulans* gene encoding an alkaline protease. Gene **150**:287–292.

132. Keil, B. 1992. Specificity of proteolysis. Springer-Verlag KG, Berlin, Germany.

133. Kikuchi, M., K. Nakahama, and K. Yoshimura. July 1985. Molecular cloning of *Bacillus amyloliquefaciens* neutral proteinase gene in *Bacillus subtilis*. Japanese patent 60,145,090.

134. Kim, M. H., S. K. Choi, B. T. Koo, B. S. Shin, C. B. Sohn, and J. Kim. 1992. Cloning and expression of *Serratia marcescens* protease gene in *Escherichia coli*. J. Microbiol. Biotechnol. **2**:231–236.

135. Kitadokora, K., E. Nakamura, M. Tamaki, T. Horii, H. Okamoto, M. Shin, T. Sato, T. Fujiwara, H. Tsuzuki, N. Yoshida, and H. Teraoka. 1993. Purification, characterization and molecular cloning of an acid amino acid-specific proteinase from *Streptomyces fradiae* ATCC 14544. Biochim. Biophys. Acta **1163**:149–157.

136. Kitano, K., S. Morita, M. Kuriyama, and K. Maejima. December 1992. Cloning, expression, and use of alkaline protease gene from *Fusarium*. European patent 519,229.

137. Kiwaki, M., H. Ikemura, M. Shimizu-Kadoba, and A. Hirashima. 1989. Molecular characterization of a cell wall-associated proteinase gene from *Streptococcus lactis* NCDO 763. Mol. Microbiol. **3**:359–369.

138. Koide, Y., A. Nakamura, T. Uozumi, and T. Beppu. 1986. Cloning and sequencing of the major intracellular serine protease gene of *Bacillus subtilis*. J. Bacteriol. **167**:110–116.

139. Kok, J., D. Hill, A. J. Haandrikman, M. Reuver, H. Laan, and G. Venema. 1988. Deletion analysis of the proteinase gene of *Streptococcus cremoris* Wg2. Appl. Environ. Microbiol. **54**:239–244.

140. Kok, J., J. M. Van Dijl, J. Van Der Vossen, and G. Venema. 1985. Cloning and expression of a *Streptococcus cremoris* proteinase in *Bacillus subtilis* and *Streptococcus lactis*. Appl. Environ. Microbiol. **50**:94–101.

141. Kok, J., K. J. Leenhouts, A. J. Haandrikman, A. M. Ledeboer, and G. Venema. 1988. Nucleotide sequence of the cell wall proteinase gene of *Streptococcus cremoris* Wg2. Appl. Environ. Microbiol. **54**:231–238.

142. Kornberg, A., J. A. Spudich, D. L. Nelson, and M. P. Deutscher. 1968. Origin of proteins in sporulation. Annu. Rev. Biochem. **37**:51–78.

143. Koszelak, S., J. D. Ng, J. Day, T. P. Ko, A. Greenwood, and A. McPherson. 1997. The crystallographic structure of the subtilisin protease from *Penicillium cyclopium*. Biochemistry **36**:6597–6604.

144. Kotler, M., R. A. Katz, and A. M. Skalka. 1988. Activity of avian retroviral protease expressed in *Escherichia coli*. J. Virol. **62**:2696–2700.

145. Kubo, M., and T. Imanaka. 1988. Cloning and nucleotide sequence of the highly thermostable neutral protease gene from *Bacillus stearothermophilus*. J. Gen. Microbiol. **134**:1883–1891.

146. Kubo, M., and Y. Higo. 1989. Expression and genetic analysis of the highly thermostable neutral protease (aspartame synthetic enzyme) gene in *Bacillus subtilis*. Toso Kenkyu Hokoku **33**:101–112.

147. Kuo, L. C., and J. A. Shafer. 1994. Retroviral proteases. Methods Enzymol. **241**:3–178.

148. Kwon, S. T., I. Terada, H. Matsuzawa, and T. Ohta. 1988. Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1 and characteristics of the deduced primary structure of the enzyme. Eur. J. Biochem. **173**:491–497.

149. Kwon, S. T., H. Matsuzawa, and T. Ohta. 1988. Determination of the positions of the disulfide bonds in aqualysin I (a thermophilic alkaline serine protease) of *Thermus aquaticus* YT-1. J. Biochem. **104**:557–559.

150. Labbe, J. P., P. Rebegrotte, and M. Turpine. 1974. Demonstrating extra-

cellular leucine aminopeptidase (EC 3.4.1.1) of *Aspergillus oryzae* (IP 410): leucine aminopeptidase 2 fraction. C. R. Acad. Sci. (Paris) **278D:**2699.

151. **Lafemina, R. L., V. V. Sardana, and C. A. Veloski.** April 1996. Stable recombinant protease from human cytomegalovirus by site-specific mutagenesis. US patent 323,953.

152. **Lao, G., and D. B. Wilson.** 1996. Cloning, sequencing and expression of a *Thermomonospora fusca* protease gene in *Streptomyces lividans.* Appl. Environ. Microbiol. **62:**4256–4259.

153. **Law, J., P. Vos., F. Hayes, C. Daly, W. de Vos, and G. Fitzgerald.** 1992. Cloning and partial sequencing of the proteinase gene complex from *Lactococcus lactis* subsp. *lactis* UC317. J. Gen. Microbiol. **138:**709–718.

154. **Lecadet, M. M., M. Lescourret, and A. Klier.** 1977. Characterization of an intracellular protease isolated from *Bacillus thuringiensis* sporulating cells and able to modify homologous RNA polymerase. Eur. J. Biochem. **79:** 329–338.

155. **Lee, C., S. Su, and R. Liaw.** 1995. Molecular analysis of an extracellular protease gene from *Vibrio parahaemolyticus.* Microbiology **141:**2569–2576.

156. **Lee, J. D., and P. E. Kolattukudy.** 1995. Molecular cloning of the cDNA and gene for an elastinolytic aspartic proteinase from *Aspergillus fumigatus* and evidence of its secretion by the fungus during invasion of the host lung. Infect. Immun. **63:**3796–3803.

157. **Lee, S., K. Yoon, H. Nam, T. Oh, S. Lee, and K. Chae.** 1996. Cloning and molecular characterization of a *nprx* gene of *B. subtilis* N515-4 encoding a neutral protease. J. Microbiol. **34:**63–73.

158. **Leenhouts, K. J., J. Gietema, J. Kok, and G. Venema.** 1991. Chromosomal stabilization of the proteinase genes in *Lactococcus lactis.* Appl. Environ. Microbiol. **57:**2568–2575.

159. **Leighton, T. J., and J. J. Stock.** 1970. Biochemical changes during fungal sporulation and spore germination. I. Phenyl methyl sulfonyl fluoride inhibition of macroconidial germination in *Microsporum gypseum.* J. Bacteriol. **101:**931–940.

160. **Leung, K. Y., and R. M. Stevenson.** 1988. Characteristics and distribution of extracellular proteases from *Aeromonas hydrophila.* J. Gen. Microbiol. **134:** 151–160.

161. **Li, S. L., S. Norioka, and F. Sakiyama.** 1990. Molecular cloning and nucleotide sequence of the B-lytic protease gene from *Achromobacter lyticus.* J. Bacteriol. **172:**6506–6511.

162. **Libby, R. T., D. Cosman, M. K. Cooney, J. E. Merriam, C. J. March, and T. P. Hopp.** 1988. Human rhinovirus 3C protease: cloning and expression of an active form in *Escherichia coli.* Biochemistry **27:**6262–6268.

163. **Lichenstein, H.** January 1993. Isolation and characterization of a novel protease from *Streptomyces lividans.* U.S. patent 724,721.

164. **Lillehoj, E. P., F. H. R. Salazar, R. J. Mervis, M. G. Raum, H. W. Chan, N. Ahmad, and S. Venkatesan.** 1988. Purification and structural characterization of the putative *gag-pol* protease of human immunodeficiency virus. J. Virol. **62:**3053–3058.

165. **Lindberg, R. A., L. D. Eirich, J. S. Price, L. Wolfinbarger, Jr., and H. Drucker.** 1981. Alkaline protease from *Neurospora crassa.* J. Biol. Chem. **256:**811–814.

166. **Liu, H., F. Muller-Plathe, and W. F. Van Gunsteren.** 1996. A combined quantum/classical molecular dynamics study of the catalytic mechanism of HIV protease. J. Mol. Biol. **261:**454–469.

167. **Liu, P. V.** 1974. Extracellular toxins of *Pseudomonas aeruginosa.* J. Infect. Dis. **130:**94–99.

168. **Liu, Y.-N., J.-L. Tang, B. R. Clark, J. M. Dow, and M. J. Danniels.** 1990. A multipurpose broad host range cloning vector and its use to characterize an extracellular protease gene of *Xanthomonas campestris* pathovar. *campestris.* Mol. Gen. Genet. **220:**433–440.

169. **Lomholt, H., K. Poulsen, and M. Kilian.** 1995. Comparative characterization of the *iga* gene encoding IgA₁ protease in *Neisseria meningitidis, N. gonorrhoeae* and *Haemophilus influenzae.* Mol. Microbiol. **15:**495–506.

170. **Lott, T. J., L. S. Page, P. Boiron, J. Benson, and E. Reiss.** 1989. Nucleotide sequence of the *Candida albicans* aspartyl proteinase gene. Nucleic Acids Res. **17:**1779.

171. **Louis, J. M., E. M. Wondrak, T. D. Copeland, C. A. Dale Smith, P. T. Mora, and S. Oroszlan.** 1989. Chemical synthesis and expression of the HIV-1 protease gene in *E. coli.* Biochem. Biophys. Res. Commun. **159:**87–94.

172. **Loutsch, J. M., N. J. Galvin, M. L. Bryant, and B. C. Holwerda.** 1994. Cloning and sequence analysis of murine cytomegalovirus protease and capsid assembly protein genes. Biochem. Biophys. Res. Commun. **203:**472–478.

173. **Lowther, W. T., P. Majer, and B. M. Duhn.** 1995. Engineering the substrate specificity of rhizopuspepsin: the role of Asp77 of fungal aspartic proteinases in facilitating the cleavage of oligopeptide substrates with lysine in P1. Protein Sci. **4:**689–702.

174. **Lu, A. Y. H., K. W. Junk, and M. J. Coon.** 1969. Resolution of the cytochrome P-450 containing W-Hydroxylation system of liver microsomes into three components. J. Biol. Chem. **244:**3714–3721.

175. **Lu, J., H. Inoue, T. Kimura, O. Makabe, and K. Takashi.** 1995. Molecular cloning of a cDNA for proctase B from *Aspergillus niger* var. *macrosporus* and sequence comparison with other aspergillopepsins I. Biosci. Biotechnol. Biochem. **59:**954–955.

176. **MacDougall, J. H., D. Beighton, and R. R. B. Russell.** 1991. Cloning and expression of protease genes from *Treponema denticola* in *Escherichia coli.* Oral Microbiol. Immunol. **6:**270–274.

177. **MacKay, V. L., S. K. Welch, M. Y. Insley, T. R. Manney, J. Holly, G. C. Saari, and M. L. Parker.** 1988. The *Saccharomyces cerevisiae* BAR1 gene encodes an exported protein with homology to pepsin. Proc. Natl. Acad. Sci. USA **85:**55–59.

178. **Manabe, K., M. Morh, M. Honjo, M. Ohoka, K. Fushimi, A. Sawakura, and Y. Furutani.** 1985. N-terminal amino acid sequences of neutral proteases from *Bacillus amyloliquefaciens* and *Bacillus subtilis:* identification of a neutral protease gene cloned in *Bacillus subtilis.* Agric. Biol. Chem. **49:**2261–2267.

179. **Masaki, T., K. Nakamura, M. Isono, and M. Soejima.** 1978. A new proteolytic enzyme from *Achromobacter lyticus* M497-1. Agric. Biol. Chem. **42:**1443–1445.

180. **Maurizi, M. R., P. Trisler, and S. Gottesman.** 1985. Insertional mutagenesis of the *lon* gene in *Escherichia coli: lon* is dispensable. J. Bacteriol. **164:**1124–1135.

181. **Maurizi, M. R., W. P. Clark, Y. Katayama, S. Rudikoff, J. Pumphrey, B. Bowers, and S. Gottesman.** 1990. Sequences and structure of ClpP, the proteolytic component of the ATP-dependent Clp protease of *Escherichia coli.* J. Biol. Chem. **265:**12536–12544.

182. **Maurizi, M. R., W. P. Clark, S.-H. Kim, and S. Gottesman.** 1990. ClpP represents a unique family of serine proteases. J. Biol. Chem. **265:**12546–12552.

183. **Menon, A. S., and A. L. Goldberg.** 1987. Protein substrates activate the ATP-dependent protease La by promoting nucleotide binding and release of bound ADP. J. Biol. Chem. **262:**14929–14934.

184. **Miller, M., M. Jasko'lski, J. K. M. Rao, J. Leis, and A. Wlodawer.** 1989. Crystal structure of a retroviral protease proves relationship to aspartic protease family. Nature **337:**576–579.

185. **Milton, D. L., A. Norqvist, and H. Wolf-Watz.** 1992. Cloning of a metalloprotease gene involved in the virulence mechanism of *Vibrio anguillarum.* J. Bacteriol. **174:**7235–7244.

186. **Mitchinson, C., and J. A. Wells.** 1989. Protein engineering of disulfide bonds in subtilisin BPN'. Biochemistry **28:**4807–4815.

187. **Miyazaki, H., N. Yanagida, S. Horinouchi, and T. Beppu.** 1989. Characterization of the precursor of *Serratia marcescens* serine protease and COOH-terminal processing of the precursor during its excretion through the outer membrane of *Escherichia coli.* J. Bacteriol. **171:**6566–6572.

188. **Mizaeno, K., T. Nakamura, T. Oshima, S. Tanaka, and H. Matsuo.** 1988. Yeast KEX2 gene encodes an endopeptidase homologous to subtilisin-like serine proteases. Biochem. Biophys. Res. Commun. **156:**246–254.

189. **Mizuno, K., and H. Matsuo.** 1984. A novel protease from yeast with specificity towards paired basic residues. Nature **309:**558–560.

190. **Moehle, C. M., M. W. Aynardi, M. R. Kolodny, F. J. Park, and E. W. Jones.** 1987. Protease B of *Saccharomyces cerevisiae:* isolation and regulation of the PRB1 structural gene. Genetics **115:**255–263.

191. **Monod, M., G. Togni, B. Hube, and D. Sanglard.** 1994. Multiplicity of genes encoding secreted aspartic proteinases in *Candida* sp. Mol. Microbiol. **13:** 357–368.

192. **Morihara, K., and K. Oda.** 1993. Microbial degradation of proteins, p. 293–364. *In* W. Guenther (ed.), Microbial degradation of natural products. VCH Publishers, Weinheim, Germany.

193. **Morita, S., M. Kuriyama, M. Nakatsu, and K. Kitano.** 1994. High level expression of *Fusarium* alkaline protease gene in *Acremonium chrysogenum.* Biosci. Biotechnol. Biochem. **58:**627–630.

194. **Moses, E. K., J. I. Rood, W. K. Yong, and G. G. Riffkin.** 1989. Molecular analysis of one of multiple protease-encoding genes from the prototype virulent strain of *Bacteroides nodosus.* Gene **77:**219–228.

195. **Murakami, K., Y. Ishida, A. Masaki, H. Tatsumi, S. Murakami, E. Nakano, H. Motai, H. Kawabe, and H. Arimura.** 1991. Isolation and characterization of the alkaline protease gene of *Aspergillus oryzae.* Agric. Biol. Chem. **55:**2807–2812.

196. **Murayama, S. Y., O. Takeda, H. Mukai, K. Takesako, E. Sano, I. Kato, and H. Yamaguchi.** 1995. *Candida albicans* aspartic proteinase: cDNA cloning and comparison among strains. Adv. Exp. Med. Biol. **362:**573–576.

197. **Nagami, Y., and T. Tanaka.** December 1986. Expression plasmid of *Bacillus natto* protease. Japanese patent 61,282,400.

198. **Nakahama, K., K. Yoshimuza, R. Maramoto, M. Kikuchi, I. Lee, T. Hase, and H. Matsubara.** 1986. Cloning and sequencing of *Serratia* protease gene. Nucleic Acids Res. **14:**5843–5855.

199. **Narhi, O. L., Y. Stabinsky, M. Levitt, L. Miller, R. Sachdev, S. Finley, S. Park, C. Kolvenbach, T. Arakawa, and M. Zukowski.** 1991. Enhanced stability of subtilisin by three point mutations. Biotechnol. Appl. Biochem. **13:**12–24.

200. **Neurath, H.** 1984. Evolution of proteolytic enzymes. Science **224:**350–357.

201. **Nicaud, J. M., E. Fabre, J. M. Beckerich, P. Fournier, and C. Gaillardin.** 1989. Cloning, sequencing, and amplification of the alkaline extracellular protease (XPR2) gene of the yeast *Yarrowia lipolytica.* J. Biotechnol. **12:** 285–297.

202. **Nielsen, T. L., S. Holmberg, and J. G. L. Petersen.** 1990. Regulated over-

production and secretion of yeast carboxypeptidase Y. Appl. Microbiol. Biotechnol. 33:307–312.

203. Nishiya, Y., and T. Imanaka. 1990. Cloning and nucleotide sequences of the *Bacillus stearothermophilus* neutral protease gene and its transcriptional activator gene. J. Bacteriol. 172:4861–4869.

204. Nissen-Meyer, J., D. Lillehaug, and I. F. Nes. 1992. The plasmid-encoded lactococcal envelope-associated proteinase is encoded by a chromosomal gene in *Lactococcus lactis* subsp. *cremoris* BC101. Appl. Environ. Microbiol. 58:750–753.

205. North, M. J. 1982. Comparative biochemistry of the proteinases of eukaryotic microorganisms. Microbiol. Rev. 46:308–340.

206. O'Day, D. H. 1976. Acid protease activity during germination of microcysts of the cellular slime mold *Polysphodylium pallidum*. J. Bacteriol. 125:8–13.

207. Ogawa, Y., H. Tatsumi, S. Murakami, Y. Ishida, K. Murakami, A. Masaki, H. Kawabe, H. Arimura, E. Nakano, H. Motai, and A. Toh-e. 1990. Secretion of *Aspergillus oryzae* alkaline protease in an osmophilic yeast, *Zygosaccharomyces rouxii*. Agric. Biol. Chem. 54:2521–2529.

208. Ohara, T., K. Makino, H. Shinagawa, A. Nakata, S. Norioka, and F. Sakiyama. 1989. Cloning, nucleotide sequence, and expression of *Achromobacter* protease I gene. J. Biol. Chem. 264:20625–20631.

209. Okada, J., H. Shimogaki, K. Murata, H. Kimura, and K. Oba. January 1986. Genetically engineered *Bacillus subtilis* for extracellular protease manufacture. Japanese patent 61,12,287.

210. Okada, Y., H. Nagase, and E. D. Harris. 1986. A metalloproteinase from human rheumatoid synovial fibroblasts that digests connective tissue matrix components. Purification and characterization. J. Biol. Chem. 261:14245–14255.

211. Osawa, M., O. Hatamoto, Y. Hanya, S. Ushijima, and K. Uchida. 1992. Cloning of alkaline protease gene of *Aspergillus*. Japanese patent 04,148,684.

212. Oshima, T., and K. Mizuno. 1989. Manufacture of KEX-2 endoproteinase and soluble derivatives using expression vectors. European patent 327,377.

213. Page, R. D. M. 1996. TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12:357–358.

214. Pantoliano, M. W., R. C. Ladner, P. N. Bryan, M. L. Rollence, J. F. Wood, and T. L. Poulos. 1987. Protein engineering of subtilisin BPN': enhanced stabilization through the introduction of two cysteines to form a disulfide bond. Biochemistry 26:2077–2082.

215. Pearl, L. H. 1987. The catalytic mechanism of aspartic proteinases. FEBS Lett. 214:8–12.

216. Pearl, L. H., and W. R. Taylor. 1987. A structural model for the retroviral proteases. Nature 329:351–354.

217. Perlmann, G. E., and L. Lorand. 1970. Proteolytic enzymes. Methods Enzymol. 19:732–735.

218. Phadatare, S., M. Rao, and V. Deshpande. 1997. A serine alkaline protease from the fungus *Conidiobolus coronatus* with a distinctly different structure than the serine protease subtilisin Carlsberg. Arch. Microbiol. 166:414–417.

219. Phadatare, S. U., M. C. Srinivasan, and V. V. Deshpande. 1993. High activity alkaline protease from *Conidiobolus coronatus* (NCL 86.8.20): enzyme production and compatibility with commercial detergents. Enzyme Microb. Technol. 15:72–76.

220. Phadatare, S. U., M. C. Srinivasan, and V. V. Deshpande. 1992. Evidence for controlled autoproteolysis of alkaline protease: a mechanism for physiological regulation of conidial discharge in *Conidiobolus coronatus*. Eur. J. Biochem. 205:679–686.

221. Phadatare, S. U., M. C. Srinivasan, and M. V. Deshpande. 1989. Evidence for the involvement of serine protease in the conidial discharge of *Conidiobolus coronatus*. Arch. Microbiol. 153:47–49.

222. Pichuantes, S., L. M. Babe, P. J. Barr, and C. S. Craik. 1989. Recombinant HIV-1 protease secreted by *Saccharomyces cerevisiae* correctly processes myristylated gag polyprotein. Proteins Struct. Funct. Genet. 6:324–337.

223. Pickersgill, R. W., I. G. Sumner, and P. W. Goodenough. 1990. Preliminary crystallographic data for protease a. Eur. J. Biochem. 190:443–444.

224. Pohlner, J., R. Halter, K. Beyreuther, and T. F. Meyer. 1986. Gene structure and extracellular secretion of *Neisseria gonorrhoeae* IgA protease. Nature 325:458–462.

225. Poldermans, B. 1990. Proteolytic enzymes, p. 108–123. *In* W. Gerhartz (ed.), Proteolytic enzymes in industry: production and applications. VCH Publishers, Weinheim, Germany.

226. Polgar, L. 1990. Common feature of the four types of protease mechanisms. Biol. Chem. Hoppe-Seyler 371:327–331.

227. Postemsky, C. J., S. S. Dignam, and P. Setlow. 1978. Isolation and characterization of *Bacillus megaterium* mutants containing decreased levels of spore protease. J. Bacteriol. 135:841–850.

228. Poulsen, K., J. Brandt, J. P. Hjorth, H. C. Thoegersen, and M. Kilian. 1989. Cloning and sequencing of the immunoglobulin A1 protease gene (*iga*) of *Haemophilus influenzae* serotype 6. Infect. Immun. 57:3097–3105.

229. Poulsen, K., J. Reinholdt, and M. Kilian. 1996. Characterization of *Streptococcus pneumoniae* immunoglobulin A1 protease gene (*iga*) and its translation product. Infect. Immun. 64:3957–3966.

230. Poulsen, K., K. Hjorth, J. Peter, and M. Kilian. 1988. Limited diversity of the immunoglobulin A1 protease gene (*iga*) among *Haemophilus influenzae*

serotype 6 strains. Infect. Immun. 56:987–992.

231. Power, M. D., P. A. Marx, M. L. Bryant, M. D. Gardner, P. J. Barr, and P. A. Luciw. 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 231:1567–1572.

232. Rahr, S., R. Halter, H. Mueller, J. Pohlner, and T. F. Meyer. 1985. Genetic analysis of neisserial immunoglobulin A1 proteases, p. 157–163. *In* G. K. Schoolnik (ed.), Pathogenic *Neisseria*. Proceedings of the Fourth International Symposium. American Society for Microbiology, Washington, D.C.

233. Ramesh, M. V., T. D. Sirakova, and P. E. Kolattukudy. 1994. Isolation, characterization and cloning of the cDNA and the gene for an elastinolytic serine proteinase from *Aspergillus flavus*. Infect. Immun. 62:79–85.

234. Ramesh, M. V., T. D. Sirakova, and P. E. Kolattukudy. 1995. Cloning and characterization of the cDNAs and genes (*mep20*) encoding homologous metalloproteinases from *Aspergillus flavus* and *Aspergillus fumigatus*. Gene 165:121–125.

235. Rao, M. B., and V. V. Deshpande. Proteases and their applications in biotechnology. *In* A. Varna (ed.), Microbes: for health, wealth and sustainable environment, in press. Malhotra Publishing House, New Delhi, India.

236. Rawlings, N. D., and A. J. Barrett. 1993. Evolutionary families of peptidases. Biochem. J. 290:205–218.

237. Reichard, U., M. Monod, and R. Ruechel. 1995. Molecular cloning and sequencing of the gene encoding an extracellular aspartic proteinase from *Aspergillus fumigatus*. FEMS Microbiol. Lett. 130:69–74.

238. Rivero, O., J. Anguita, C. Paniagua, and G. Naharro. 1990. Molecular cloning and characterization of an extracellular protease gene from *Aeromonas hydrophila*. J. Bacteriol. 172:3905–3908.

239. Rivero, O., J. Anguita, D. Mateos, C. Paniagua, and G. Naharro. 1991. Cloning and characterization of an extracellular temperature-labile serine protease gene from *Aeromonas hydrophila*. FEMS Microbiol. Lett. 81:1–8.

240. Roberts, B. 1989. Nuclear location signal-mediated protein transport. Biochim. Biophys. Acta 1008:263–280.

241. Roberts, J. W., C. W. Roberts, and D. W. Mount. 1977. Inactivation and proteolytic cleavage of phage repressor *in vitro* in an ATP-dependent reaction. Proc. Natl. Acad. Sci. USA 74:2283–2287.

242. Rougeau, C. T., M. W. Lauvel, and H. Fukuhara. 1988. The *Kluyveromyces lactis* KEX1 gene encodes a subtilisin-type serine protease. FEBS Lett. 234:464–470.

243. Russel, A. J., and A. R. Ferhst. 1987. Rational modification of enzyme catalysis by engineering surface charge. Nature 328:496–500.

244. Russel, A. J., P. G. Thomas, and A. R. Fersht. 1987. Electrostatic effects on modification of charged groups in the active site cleft of subtilisin by protein engineering. J. Mol. Biol. 193:803–813.

245. Sagata, N., T. Yasunaga, Y. Ogawa, J. Tsuzuku-Kawamara, and Y. Ikawa. 1984. Bovine leukemia virus. Unique structural features of its long terminal repeats and evolutionary relationship to human T-cell leukemia virus. Proc. Natl. Acad. Sci. USA 81:4741–4745.

246. Saiga, A., T. Tanaka, S. Orita, A. Sato, S. Sato, T. Hachisu, K. Abe, Y. Kimura, Y. Kondo, et al. 1993. Human T-cell leukemia virus type I protease protein expressed in *Escherichia coli* possesses aspartic proteinase activity. Arch. Virol. 128:195–210.

247. Samal, B., B. Karan, T. C. Boone, K. K. Chen, M. F. Rahde, and Y. Stabinsky. 1989. Cloning and expression of the gene encoding a novel proteinase from *Tritirachium album* Limber. Gene 85:329–333.

248. Sanchez-Pescador, R., M. D. Power, P. J. Borr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, A. Renar, A. Randoph, J. A. Levy, D. Dina, and P. A. Luciw. 1985. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science 227:484–492.

249. Saul, D. J., L. C. Williams, H. S. Toogood, R. M. Daniel, and P. L. Bergquist. 1996. Sequence of the gene encoding a highly thermostable neutral proteinase from *Bacillus* sp. strain EA1: expression in *Escherichia coli* and characterization. Biochim. Biophys. Acta 1308:74–80.

250. Schechter, I., and A. Berger. 1967. On the size of the active site in proteases I papain. Biochem. Biophys. Res. Commun. 27:157–162.

251. Schmidt, B. F., L. Woodhouse, R. M. Adams, T. Ward, S. E. Mainzer, and P. J. Lad. 1995. Alkalophilic *Bacillus* sp. strain LG12 has a series of serine protease genes. Appl. Environ. Microbiol. 61:4490–4493.

252. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. Cell 32:853–869.

253. Shannon, J. D., E. N. Baramova, J. B. Bjarnason, and J. W. Fox. 1989. Amino acid sequence of a *Crotalus atrox* venom metalloprotease which cleaves type IV collagen and gelatin. J. Biol. Chem. 264:11575–11583.

254. Shibano, Y., K. Morihara, K. Okuda, and A. Fukushima. December 1991. Molecular cloning of alkaline protease gene of *Pseudomonas aeruginosa*. Japanese patent 03,277,278.

255. Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura. 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA 82:3101–3105.

256. Shinick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature 293:543–548.

257. Shintani, T., and E. Ichishima. 1994. Primary structure of aspergillopepsin

I deduced from nucleotide sequence of the gene and aspartic acid-76 is an essential active site of the enzyme for trypsinogen activation. Biochim. Biophys. Acta 1204:257–264.

258. Sidorenkov, I. N., E. A. Nosovskaya, and S. V. Kostrov. 1995. Cloning and expression of Bacillus stearothermophilus-thermostable metallo-protease gene in Bacillus subtilis. Mol. Biol. 29:507–511.

259. Sielecki, A. R., M. Fujinaga, R. J. Read, and M. N. G. James. 1991. Refined structure of porcine pepsinogen at 1.8A resolution. J. Mol. Biol. 219:671–692.

260. Silen, J. L., C. N. McGrath, K. R. Smith, and D. A. Agard. 1988. Molecular analysis of the gene encoding α-lytic protease: evidence for a preproenzyme. Gene 69:237–244.

261. Silen, J. L., D. Frank, A. Fujishige, R. Bone, and D. A. Agard. 1989. Analysis of prepro-α-lytic proteases expression in Escherichia coli reveals that the proregion is required for activity. J. Bacteriol. 171:1320–1325.

262. Sirakova, T. D., A. Markaryan, and P. E. Kolattukudy. 1994. Molecular cloning and sequencing of the cDNA and gene for a novel elastinolytic metalloproteinase from Aspergillus fumigatus and its expression in Escherichia coli. Infect. Immun. 62:4208–4218.

263. Sloma, A., A. Ally, D. Ally, and J. Pero. 1988. Gene encoding a minor extracellular protease in Bacillus subtilis. J. Bacteriol. 170:5557–5563.

264. Sloma, A., C. F. Rudolph, G. A. Rufo Jr., B. J. Sullivan, K. A. Theriault, D. Ally, and J. Pero. 1990. Gene encoding a novel extracellular metalloprotease in Bacillus subtilis. J. Bacteriol. 172:1024–1029.

265. Sloma, A., G. A. Rufo, Jr., C. F. Rudolph, B. J. Sullivan, K. A. Theriault, and J. Pero. 1990. Bacillopeptidase F of Bacillus subtilis: purification of the protein and cloning of the gene. J. Bacteriol. 172:1470–1477.

266. Smeekens, S. P. 1993. Processing of protein precursors by a novel family of subtilisin-related mammalian endoproteases. Bio/Technology 11:182–186.

267. Sobek, H., H. J. Hecht, B. Hoffman, W. Aeble, and D. Schomburg. 1990. Crystal structure of an alkaline protease from Bacillus alkalophilus at 2.4 A resolution. FEBS Lett. 274:57–66.

268. Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson. 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45:375–385.

269. Stabile, M. R., W. G. Lai, G. DeSantis, M. Gold, and J. B. Jones. 1996. Probing the specificity of the S₁ binding site of M222 mutants of subtilisin B. lentus with boronic acid inhibitors. Bioorg. Med. Chem. Lett. 6:2501–2506.

270. Stahl, M. L., and E. Ferrari. 1984. Replacement of the Bacillus subtilis subtilisin structural gene with an in vitro-derived deletion mutation. J. Bacteriol. 158:411–418.

271. Steffy, K. R., S. Schoen, and C. Chen. 1995. Nucleotide sequence of the herpes simplex virus type 2 gene encoding the protease and capsid protein ICP35. J. Gen. Virol. 76:1069–1072.

272. Stoll, E., H. G. Weder, and H. Zuber. 1976. Aminopeptidase II from Bacillus sterothermophilus. Biochim. Biophys. Acta 438:212–220.

273. Strausberg, S. L., P. A. Alexander, D. T. Gallagher, G. L. Gilliland, B. L. Barnett, and P. B. Bryan. 1995. Directed evolution of a subtilisin with calcium-independent stability. Bio/Technology 13:669–673.

274. Sussman, M. D., and P. Setlow. 1991. Cloning, nucleotide sequence and regulation of the Bacillus subtilis gpr gene, which codes for the protease that initiates degradation of small, acid-soluble proteins during spore germination. J. Bacteriol. 173:291–300.

275. Suzuki, C. K., M. Rep, J. M. Van Dijl, K. Suda, L. A. Grivell, and G. Schatz. 1997. ATP-dependent proteases that also chaperone protein biogenesis. Trends Biochem. Sci. 22:118–123.

276. Suzuki, T., A. Itoh, S. Ichihara, and S. Mizushima. 1987. Characterization of the sppA gene coding for protease IV, a signal peptide peptidase of Escherichia coli. J. Bacteriol. 169:2523–2528.

277. Svendsen, I., M. R. Jensen, and K. Breddman. 1991. The primary structure of the glutamic acid specific protease of Streptomyces griseus. FEBS Lett. 292:165–167.

278. Takagi, H., T. Morinaga, H. Ikemuar, and M. Inouye. 1989. The role of pro-239 in the catalysis and heat stability of subtilisin E. Eur. J. Biochem. 105:953–956.

279. Takagi, M., T. Imanaka, and S. Aiba. 1985. Nucleotide sequence and promoter region for the neutral protease gene from Bacillus stearothermophilus. J. Bacteriol. 163:824–831.

280. Takagi, H., T. Takahashi, H. Momose, M. Inouye, Y. Maeda, H. Matsuzawa, and T. Ohta. 1990. Enhancement of the thermostability of subtilisin E by introduction of a disulfide bond engineered on the basis of structural comparison with a thermophilic serine protease. J. Biol. Chem. 265:6874–6878.

281. Takagi, H., Y. Morinaga, H. Ikemura, and M. Inouye. 1988. Mutant subtilisin E with enhanced protease activity obtained by site directed mutagenesis. J. Biol. Chem. 263:19592–19596.

282. Takahashi, K. 1987. The amino acid sequence of rhizopuspepsin, an aspartic proteinase from Rhizopus chinensis. J. Biol. Chem. 262:1468–1478.

283. Takahashi, K., H. Inoue, T. Kimura, C. Nojiri, and O. Makabe. March 1992. Cloning and expression of acid protease A gene of Aspergillus. Japanese patent 04,75,586.

284. Takeuchi, M., K. Ogura, T. Hamamoto, and Y. Kobayashi. 1995. Molecular cloning and sequence analysis of a gene encoding an aspartic proteinase from Aspergillus oryzae. Adv. Exp. Med. Biol. 362:577–580.

285. Tanguy-Rangean, C., M. Wesolowski-Louvel, and H. Fukuhare. 1988. The Kluyveromyces lactis KEX-1 gene encodes a subtilisin-type serine proteinase. FEBS Lett. 234:464–470.

286. Tatsumi, H., M. Ohsawa, R. F. Tsuji, S. Murakami, E. Nakano, H. Motai, A. Masaki, Y. Ishida, K. Murakami, H. Kawabe, and H. Arimura. 1988. Cloning and sequencing of the alkaline protease cDNA from Aspergillus oryzae. Agric. Biol. Chem. 52:1887–1888.

287. Tatsumi, H., S. Murakami, Y. Ogawa, A. Masaki, Y. Ishida, K. Murakami, H. Kuwabe, H. Arimura, E. Nakano, and H. Motai. 1991. Autoproteolytic processing and processing dependent secretion of Aspergillus oryzae protease in yeast. Agric. Biol. Chem. 55:3099–3101.

288. Tatsumi, H., Y. Ogawa, S. Murakami, Y. Ishida, K. Murakami, A. Masaki, H. Kawabe, H. Arimura, E. Nakano, and H. Motai. 1989. A full length cDNA clone for the alkaline protease from Aspergillus oryzae: structural analysis and expression in Saccharomyces cerevisiae. Mol. Gen. Genet. 219:33–38.

289. Tawara, T., S. Nanba, and J. Nagai. April 1989. Alkaline proteinase, its cloning and recombinant manufacture with Bacillus. Japanese patent 01,101,886.

290. Thomas, P. G., A. J. Russell, and A. R. Fersht. 1985. Tailoring the pH dependence of enzyme catalysis using protein engineering. Nature 28:375–376.

291. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. Clustal W: improving sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. 22:4673–4680.

292. Tigue, N. J., P. J. Matharu, N. A. Roberts, J. S. Mills, J. Kay, and R. Jupp. 1996. Cloning, expression and characterization of the proteinase from human herpesvirus 6. J. Virol. 70:4136–4141.

293. Tobe, S., M. Odera, and Y. Asai. July 1992. Cloning and expression of gene for alkaline protease Ya of Bacillus. Japanese patent 04,197,182.

294. Togni, G., D. Sanglard, R. Falchetto, and M. Monod. 1991. Isolation and nucleotide sequence of the extracellular acid protease gene (ACP) from yeast Candida tropicalis. FEBS Lett. 286:181–185.

295. Toma, S., M. Delbue, A. Pirola, and G. Grandi. 1986. nprR1 and nprR2 regulatory regions for neutral protease expression in Bacillus subtilis. J. Bacteriol. 167:740–743.

296. Tonouchi, N., H. Shoun, T. Uozumi, and T. Beppu. 1986. Cloning and sequencing of a gene for Mucor rennin, an aspartate protease from Mucor pusillus. Nucleic Acids Res. 14:7557–7568.

297. Tran, L., X. Wu, and S. Wong. 1991. Cloning and expression of a novel protease gene encoding an extracellular neutral protease from Bacillus subtilis. J. Bacteriol. 173:6364–6372.

298. Tsujibo, H., K. Miyamoto, K. Tanaka, M. Kawai, T. Tainaka, C. Imada, Y. Okami, and Y. Inamori. 1993. Cloning and sequence of an alkaline serine protease-encoding gene from the marine bacterium Alteromonas sp. strain O-7. Gene 136:247–251.

299. Tsujibo, H., K. Miyamoto, K. Tanaka, Y. Kaidzu, C. Imada, Y. Okami, and Y. Inamori. 1996. Cloning and sequence analysis of a protease encoding gene from the marine bacterium Alteromonas sp. strain O-7. Biosci. Biotechnol. Biochem. 60:1284–1288.

300. Van der Laan, J. C., G. Gerritse, L. Mulleners, R. Van der Hoek, and W. J. Quax. 1991. Cloning, characterization and multiple chromosomal integration of a Bacillus alkaline protease gene. Appl. Environ. Microbiol. 57:901–909.

301. Vanderslice, P., C. S. Craik, J. A. Nadel, and G. H. Caughey. 1989. Molecular cloning of dog mast cell tryptase and a related protease: structural evidence of a unique mode of serine protease activation. Biochemistry 28:4148–4155.

302. Vanderslice, P., S. M. Ballinger, E. K. Tam, S. M. Goldstein, C. S. Kraik, and G. H. Caughey. 1990. Human mast cell tryptase: multiple cDNAs and genes reveal a multigene serine protease family. Proc. Natl. Acad. Sci. USA 87:3811–3815.

303. Van Melderen, L., M. Thi, P. Leechi, S. Gottesman, M. Couturier, and M. L. Maurizi. 1996. ATP-dependent degradation of Ccd A by Lon protease. J. Biol. Chem. 271:27730–27738.

304. Vasantha, N., L. D. Thompson, C. Rhodes, C. Banner, J. Nagle, and D. Filpula. 1984. Genes for alkaline protease and neutral protease from Bacillus amyloliquefaciens contain a large open reading frame between the regions coding for signal sequence and mature protein. J. Bacteriol. 159:811–819.

305. Visser, S., F. A. Exterkate, C. J. Slangen, and G. J. De Veer. 1986. Comparative study of action of cell wall proteinases from various strains of Streptococcus cremoris on bovine α- and κ-casein. Appl. Environ. Microbiol. 52:1162–1166.

306. Vos, H. L., F. M. Van der Lee, A. Reemst, A. E. Van Loon, and J. S. Sussenbach. 1988. The genes encoding the DNA binding protein and the 23 K protease of adenovirus type 40 and 41. Virology 163:1–10.

307. Wandersman, C., P. Delepelaire, S. Letoffe, and M. Schwartz. 1987. Char-

acterization of *Erwinia chrysanthemi* extracellular proteases: cloning and expression of the protease genes in *Escherichia coli*. J. Bacteriol. **169**:5046–5053.

308. **Wani, J. H., J. V. Gilbert, A. G. Plant, and J. N. Weiser.** 1996. Identification, cloning and sequencing of the immunoglobulin A1 protease gene of *Streptococcus pneumoniae*. Infect. Immun. **64**:3967–3974.

309. **Ward, O. P.** 1983. Proteinases, p. 251–305. *In* W. M. Fogarty (ed.), Microbial enzymes and biotechnology. Applied Science Publishers, London, England.

310. **Watson, R. R.** 1976. Substrate specificities of aminopeptidases: a specific method for microbial differentiation. Methods Microbiol. **9**:1–14.

311. **Weaver, L. H., W. R. Kester, and B. W. Matthews.** 1977. A crystallographic study of the complex of phosphoramidon with thermolysin. A model for the presumed catalytic transition state and for the binding of structures. J. Mol. Biol. **114**:119–132.

312. **Wells, J. A., and D. B. Powers.** 1986. *In vivo* formation and stability of engineered disulfide bonds in subtilisin. J. Biol. Chem. **261**:6564–6570.

313. **Wells, J. M., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen.** 1983. Cloning, sequencing and secretion of *Bacillus amyloliquefaciens* subtilisin in *Bacillus subtilis*. Nucleic Acids Res. **11**:7911–7925.

314. **Wilhelm, S. M., I. E. Collier, A. Kronberger, A. Z. Eisen, B. L. Marmer, G. A. Grant, E. A. Bauer, and G. I. Goldberg.** 1987. Human skin fibroblast stromelysin: structure, glycosylation, substrate specificity and differential expression in normal and tumorigenic cells. Proc. Natl. Acad. Sci. USA **84**:6725–6729.

315. **Woolford, C. A., L. B. Daniels, F. J. Park, E. W. Jones, J. N. Van Arsdell, and M. A. Innis.** 1989. The PEP4 gene encodes an aspartyl protease implicated in the post-translational regulation of *Saccharomyces cerevisiae* vacuolar hydrolases. Mol. Cell. Biol. **6**:2500–2510.

316. **Wu, M., C. Y. Yeong, S. C. Lee, and S. M. Wong.** 1993. Nucleotide sequence of the 3′ half of zucchini yellow mosaic virus (Singapore isolate) genome encoding the 4 k protein, protease, polymerase and coat protein. Nucleic Acids Res. **21**:1317.

317. **Xu, F.-F., L. E. Pearce, and P.-L. Yu.** 1990. Molecular cloning and expression of a proteinase gene from *Lactococcus lactis* subsp. *cremoris* H2 and construction of a new lactococcal vector pFX1. Arch. Microbiol. **154**:99–104.

318. **Yamagata, Y., and E. Ichishima.** 1995. A new alkaline serine protease from alkalophilic *Bacillus* sp.: cloning, sequencing and characterization of an intracellular protease. Curr. Microbiol. **30**:357–366.

319. **Yamagata, Y., R. Abe, Y. Fujita, and E. Ichishima.** 1995. Molecular cloning and nucleotide sequence of the 90K serine protease gene, hsp K, from *Bacillus subtilis* (natto) no. 16. Curr. Microbiol. **31**:340–344.

320. **Yamashita, I., D. Hirata, M. Machida, and S. Fukui.** 1986. Cloning and expression in *Saccharomyces cerevisiae* of the secretable acid protease gene from *Saccharomycopsis fibuligera*. Agric. Biol. Chem. **50**:109–113.

321. **Yamashita, T., N. Tonouchi, T. Vozami, and T. Beppu.** 1987. Secretion of mucor rennin, a fungal aspartic protease of *Mucor pusillus*, by recombinant yeast cells. Mol. Gen. Genet. **210**:462–467.

322. **Yanagida, N., T. Vozumi, and T. Beppu.** 1986. Specific excretion of *Serratia marcescens* protease through the outer membrane of *Escherichia coli*. J. Bacteriol. **166**:937–944.

323. **Yang, M. Y., E. Ferrari, and D. J. Henner.** 1984. Cloning of the neutral protease gene of *Bacillus subtilis* and the use of the cloned gene to create an in vitro-derived deletion mutation. J. Bacteriol. **160**:15–21.

324. **Yang, Q., X. Jiang, L. Wu, Y. Cui, and S. Yang.** 1991. Cloning of a thermostable neutral protease gene from *Bacillus stearothermophilus* in *Bacillus subtilis*. Shengwu Gongcheng Xuebao **7**:207–212.

325. **Yang, S., C. Jin, D. Bao, and S. Zhang.** 1996. Cloning and expression of thermostable protease gene from *Bacillus stearothermophilus* HY-69 in *Bacillus subtilis*. Ann N. Y. Acad. Sci. **782**:219–225.

326. **Yoneda, Y., N. Imamoto-Sonobe, Y. Matsuoka, R. Iwamoto, Y. Kiho, and T. Uchida.** 1988. Antibodies to Asp-Asp-Glu-Asp can inhibit transport of nuclear proteins into the nucleus. Science **242**:275–278.

327. **Yoshimoto, T., H. Oyama, T. Honda, H. Tone, T. Takeshita, T. Kamiyama, and D. Tsuru.** 1988. Cloning and expression of subtilisin amylosacchariticus gene. J. Biochem. **103**:1060–1065.

328. **Yoshimoto, T., H. Oyama, T. Takeshita, H. Higashi, S. L. Xu, and D. Tsura.** 1990. Nucleotide sequence of the neutral protease gene from *Bacillus subtilis* var. *amylosacchariticus*. J. Ferment. Bioeng. **70**:370–375.

329. **Yoshimoto, T., R. Walter, and D. Tsuru.** 1980. Proline-specific endopeptidase from *Flavobacterium*: purification and properties. J. Biol. Chem. **225**:4786–4792.

330. **Yoshimura, K., T. Miyazaki, K. Nakahama, and M. Kikuchi.** 1985. Cloning and partial nucleotide sequence of a neutral protease gene from *Bacillus amyloliquefaciens*. Takeda Kenkyusho Ho **44**:42–50.

331. **Young, T. W., A. Wadeson, D. J. Glover, R. V. Quincey, M. J. Butlin, and E. A. Kamei.** 1996. The extracellular acid protease gene of *Yarrowia lipolytica*: sequences and pH regulated transcription. Microbiology **142**:2913–2921.

332. **Young, T. W., and E. A. Hosford.** 1987. Genetic manipulation of *Saccharomyces cerevisiae* to produce extracellular protease, p. 521–528. *In* Proceedings of the 21st Congr. -European Brewers Convention.

333. **Yu, F., H. Yamada, K. Daishima, and S. Mizushima.** 1983. Nucleotide sequence of the *lspA* gene, the structural gene for lipoprotein signal peptidase of *Escherichia coli*. FEBS Lett. **173**:264–268.

334. **Zehnbauer, B. A., and A. Markovitz.** 1980. Cloning of gene *lon* (CapR) in *Escherichia coli* K-12 and identification of polypeptide specified by the cloned deoxyribonucleic acid fragment. J. Bacteriol. **143**:852–863.

335. **Zen-Yoji, H., T. Terayama, and M. Benoshi.** 1961. Studies on staphylococcal coagulase I. Antigenic difference of coagulase and distribution of the anticoagulase in human sera. Jpn. J. Microbiol. **5**:237–247.

336. **Zwickl, P., A. Grziwa, G. Puhler, B. Dahlmann, F. Lottspeich, and W. Baumeister.** 1992. Primary structure of the *Thermoplasma* proteasome and its implications for the structure, function and evolution of the multicatalytic proteinase. Biochemistry **31**:964–972.

337. **Zwickl, P., F. Lottspeich, B. Dahlmann, and W. Baumeister.** 1991. Cloning and sequencing of the gene encoding the large (α-) subunit of the proteasome from *Thermoplasma acidophilum*. FEBS Lett. **278**:217–221.
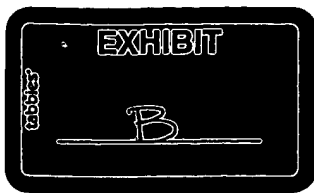
ORIGINAL PAPER

Isolation and Characterization of Two Cathepsins from Muscle of *Carassius auratus gibelio.*

D. DINU*, I. F. DUMITRU, M. T. NECHIFOR

*Faculty of Biology, University of Bucharest, 91-95 Spl. Independentei, 76201 Bucharest, Romania

Abstract

Two cathepsins were identified in the white skeletal muscle from Carassius auratus gibelio. One of these, cathepsin D, has optimal activity at pH 3.5 with hemoglobin as substrate and a molecular weight 38,200 Da. The second, cathepsin E, is a protein with a molecular weight of 82,000 Da and an optimum pH 2.5. Both of the enzymes were strongly inhibited by pepstatin, a specific inhibitor for aspartic proteinases.

Introduction

Proteolytic cleavage of peptide bonds is one the most frequent and important enzymatic modifications of proteins. Proteolytic processing is the final step in the expression of activity of a great variety of proteins. Processing occurs in many different ways and is triggered by many different kinds of proteases. However, in every known case, proteolysis is directed and limited to the cleavage of specific peptide bonds in the target protein. The key to this selectivity is limited proteolysis, which depends on the accessibility of the scissile peptide bond to the processing protease and on its specificity. Compact protein domains are usually resistant to proteolysis, in contrast to more flexible surface loops and interdomain regions that can adapt themselves to the active site of the protease [1].

Proteolytic enzymes are ubiquitously distributed in all biological tissues and fluids. The best characterized are mammalian digestive proteases such as pepsin, trypsin, chymotrypsin and elastase. The digestive proteases are involved in the hydrolysis of dietary proteins and do not play a role in protein turnover within an organism. Much less is known, by comparison, about intracellular tissue proteases, their enzymatic specificity and physiological substrates [2]. The repertoire of proteases that are integral components of cells is enormous and in part unexplored. To mention a few, it includes the entire class of lysosomal proteases (cathepsins), membrane-bound proteases and proteases of specialized tissues such as the reproductive tracts, muscle, skin, lens and adrenals.

The purpose of the present study was to investigate the proteolytic enzyme from white skeletal muscle of *Carassius auratus gibelio.* Many kinds of proteolytic enzymes were reported in fish. A human-like cathepsin D protease was isolated from the salmon *(Oncorhynchus masou)* ovary [3]. A caspase, which plays an important role in apoptosis in

fish cells and zebrafish (*Danio rerio*) embryo, was reported [4]. From chum salmon (*Oncorhynchus Keta*), an anionic trypsin acting on p- aminophenyl ester, was isolated [5]. In rainbow trout matrix was detected a metalloproteinase capable of degrading type I collagen [6].

Materials and Methods

Enzyme's extraction

After the fish killed, the white skeletal muscle was quickly dissected and rinsed with cold saline solution to remove blood. The tissue was suspended in demineralised water and was disrupted with a Polytron homogenizer at 0-4°C. The disrupted cells were submitted at a subcellular fractionation by differential centrifugation. In this procedure the homogenate was centrifuged at 600 g for 10 minutes, than the supernatant obtained was centrifuged for 10 minutes at 10,000 g. The pellets were suspended in 0,1 M TRIS-HCl buffer, pH 7.4 and centrifuged at 4,000 g. The clear supernatant obtained continents mainly proteins from mitochondria and lysosomes and was utilized for investigation.

Assay of proteolytic activity

The proteolytic activity (cathepsin activity) was assayed according to Barret [7], using acid-denatured hemoglobin as a substrate. A 8% hemoglobin solution was diluted about 3-fold with distilled water, then acidified to pH 2.0 with 1 N HCl, and the final concentration of hemoglobin was brought to 2% with distilled water. One milliliter of the substrate solution is incubated with an appropriate amount of enzyme solution at 37°C. The reaction is stopped by addition of 2 ml of 5% trichloroacetic acid. The mixture was filtered and the soluble reaction products were treated with Folin reagent and measured at 750nm.

Units of activity

There is no international agreement about units obtained by the hemoglobin-digestion method. Some investigators, including Anson and Mirsky [8] used the amount of tyrosine-equivalent in trichloroacetic acid-soluble peptide. In the current study, one unit of proteolytic activity (U) was defined by the amount of enzyme will remove 1μmol of tyrosine in a minute from the acid-denatured hemoglobin.

Determination of protein

Protein concentration was determined using the Bradford method [9], with bovine serum albumin as standard, and, during the course of enzyme purification, by measurement of A_{280}.

Testing of potential inhibitors

Portions of the enzyme solution were mixed with potential inhibitors and incubated for 30 minutes at 37°C. Mixtures were then incubated with acid-denatured hemoglobin. Reagents blanks were also run for each potential inhibitor. The percentage inhibition was determined by comparing the activities with those measured for positive controls that contained no inhibitor.

Results and Discussions

The crude extract of white muscle of *Carassius auratus gibelio* was fractionated with acetone. The 40 − 70% acetone precipitate was collected by centrifugation at 15,000 g for 15 minutes at −5°C. This was redissolved in 9% NaCl and applied on a Bio-Gel P-100 column (1.6 x 72 cm) preequilibrated with the same solution. The column was run at 10 ml/h with a solution 9% of NaCl. The typical elution profile is shown in **Figure 1**. Proteolytic activity against hemoglobin was detected as two peaks. We designated these two peaks as cathepsin I

and cathepsin II because the cathepsins are the only proteolytic enzymes founded in lysosomrs.
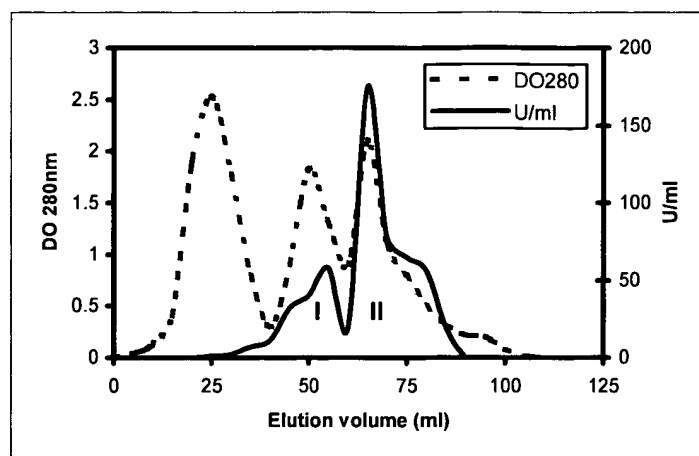
Figure 1. Separation of cathepsins by gel-filtration on Bio-Gel P-100.

The molecular weights of the two proteolytic enzymes were estimated using a calibration kit with: aprotin (6,500), citocrom c (12,400), carbonic anhydrase (29,000), albumin (66,000) and alcohol dehydrogenase (150,000). The elution volume of cathepsin I matched a molecular weight of 82,000 Da. For the cathepsin II, the molecular weight was estimated to 38,200 Da (**Figure 2**).
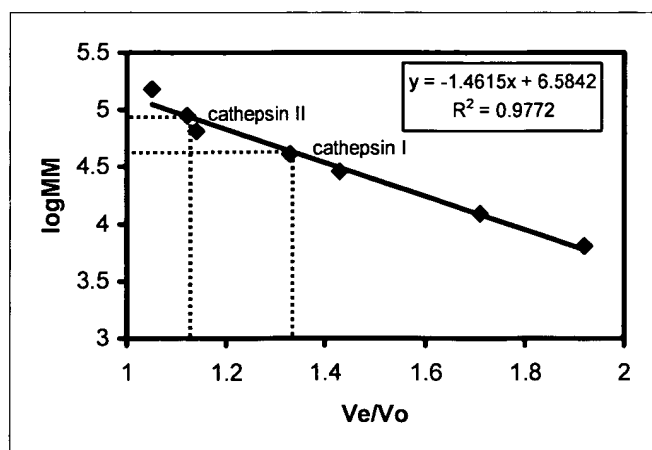
Figure 2. Molecular weights of cathepsins.

The activities of both proteolytic enzymes were higher of acid pH. The cathepsin I hydrolyzed hemoglobin most rapidly at around 2.5 and at a significantly lower rate at higher pH values such as pH 4.0 and pH 5.0. For the enzyme II the maximal proteolytic activity was at pH 3.5 (**Figure 3**).
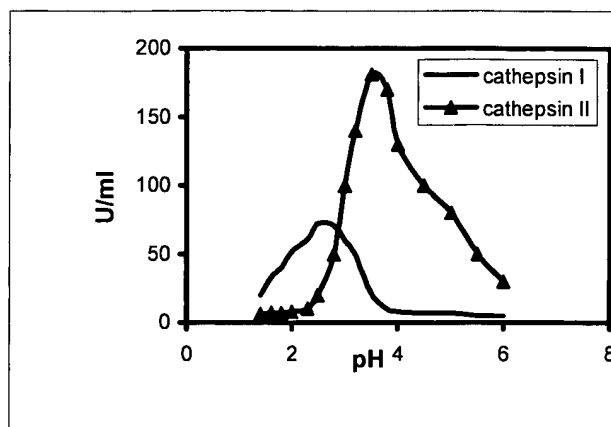
Figure 3. Hydrolytic activities of the cathepsins at different pH values.

The both proteolytic enzymes had hydrolytic activity on various protein substrates. Protein substrates such as hemoglobin, serum albumin, and casein were hydrolyzed efficiently (**Table 1**). Hemoglobin was the best protein substrate for both enzymes.

Table 1. Hydrolyses of proteins substrates by cathepsins I and II.

Substrate	Relative activity* (%)	
	Cathepsin I	Cathepsin II
Hemoglobin	100	100
Bovine serum albumin	22.3	52.3
Ovalbumin	0	2,5
γ-globulin	5.2	7.4
Casein	30.2	23.6

* Activity with hemoglobin as substrate is taken as 1005 in each case.

Potentials inhibitors of proteolytic activities were tested and the results are presented in **Table 2**. Iodoacetic acid and iodoacetic amide, compounds that irreversibly inactivate cysteine proteinases, did not significally alter the activity of both cathepsins from white skeletal muscle from *Carassius auratus gibelio*. Diisopropyl phosphofluoridate and phenylmethanesulphonyl fluoride, potent inhibitors of serine proteinases had no effects on cathepsins. The both proteolytic activities were inhibited only by pepstatin, one of the most specific inhibitor in enzymology that is highly selective for the aspartic proteinases.

Table 2. Effect of potential inhibitors on cathepsin I and II.

Compound	Final concentration	Inhibition (%)	
		Cathepsin I	Cathepsin II
Iodoacetic acid	10 mM	5.3	4.5
Iodoacetic amide	10 mM	4.1	4.6
Phenylmethanesulphonyl fluoride	1 mM	6.2	3.4
Diisopropyl phosphofluoridate	1 mM	2.5	3.5
Pepstatin	1 μM	98.2	97.5

The only intracellular proteinases that may be inhibited by pepstatin are cathepsins D (EC 3.4.23.5) and cathepsins E (EC 3.4.23.34 [10]. These two cathepsins are specificity similar to pepsin. Hemoglobin is the best protein substrate for these enzymes, and its hydrolysis proceeds most rapidly at very low pH values.

The proprieties of cathepsin I, pH optimum 2.5 and molecular weight 82,000 Da, suggest that it is a cathepsin E. Catepsins E hydrolysis hemoglobin most rapidly at around pH 2.5 at a significantly lower rate at higher pH values such as pH 4.0 and pH 5.0 [11] and has a molecular weight about 76,000 – 80,000 Da [12]. The cathepsin II seems to be a cathepsin D, enzyme with a molecular weight about 38,000 – 50,000 Da and an optimum pH 3.5 [11].

The properties of cathepsin D isolated from white skeletal muscle of Carassius *auratus gibelio* are similar to those of other fish catepsins. A cathepsin D with a molecular weight of 38,000-39,000 Da, an optimum pH 2.5 and p 6.8 was purified from herring muscle (*Clupea harengus*).From herring muscle (*Clupea harengus*) was purified a cathepsin D with a molecular weight 38,000-39,000 Da, an optimum pH 2.5 and a pI 6.8. It was inhibited by pepstatin and it was able to degrade myosin, actin and tropomyosin [13]. Sex- and tissue-specific expression of aspartic proteinases was shown in zebrafish (*Dario rerio*) [14]. Antibacterial cathepsins in different types of ambicoloured Japanase flounder skin were reported [15].

References

1. NEURATH H, The diversity of proteolytic enzymes. In: Beynon R.J., Bond J.S., eds., Proteolytic enzymes, IRL Press, Oxford University Press, pp. 1-12, 1990.

2. BOND J.S., BUTLER P.E., *Ann.Rev. Biochem.,* **56**: 333, 1987.

3. HIRAMATSU N., ISCHIKAWA N., FUKADA H., FUJITA T., SULIVAN C.V., HARA A., *J. Exp. Zool.,* **292**: 11, 2002.

4. YABU T., KISHI O., OKAZAKI T., YAMASHITO M., *Biochem. J.,* **360(part 1)**: 39, 2001.

5. SEKIZAKI H., ITOH K., MURAKAMI M., TOYOTA E., TANIZAWA K., *Comp. Biochem. Physiol.B. Biochem. Mol. Biol.,* **127**: 337, 2000.

6. SAITO M., SATO K., KUMISAKI N., KIMURA S., *Eur. J. Biochem.,* **267**: 6943, 2000.

7. BARETT A.J., *Biochem. J.,* **117**: 601, 1970.

8. ANSON M.L., Mirsky A.E., *J. Gen. Physiol.,* **16**: 59, 1932.

9. BRADFORD M. M., *Anal. Biochem.,* **72**: 248, 1976.

10. SALVENSEN G., NASAGE H., Inhibition of proteolytic enzymes. In: Beynon R.J., Bond J.S., eds., Proteolytic enzymes, IRL Press, Oxford University Press, pp. 83-102, 1990.

11. FITZGERALD P. M. D., Mc KEENEN B. M., VAN MIDDLESWORTH J. F., SPRINGER J. P, HEIMACH J. C., LEU C. T., DIXON A. F., DARKE P. L., *J. Biol. Chem.,* **265**: 14209, 1990.

12. KAGEYAMA T., Procathepsin E and cathepsin E. In: Barrett A.J., ed., Methods in Enzymology, pp. 120-136, 1995.

13. NIELSON L. B., NIELSEN H. H., *Comp. Biochem. Physiol.B. Biochem. Mol. Biol.,* **128**: 351, 2001.

14. RIGGIO M., SCUDIERO R., FILOSA S., PARISI E., *Gene,* **360**: 67, 2000.

15. ARANISHI F., MANO N., *Fish Shellfish Immunol.* **10**: 87, 2000.

758

Roum. Biotechnol. Lett., Vol. 7, No. 3, 753-758 (2002)
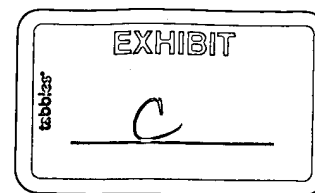
The Inhibition of Tissue Acid Proteinases by Pepstatin

By A. J. BARRETT and J. T. DINGLE
Tissue Physiology Department, Strangeways Research Laboratory, Cambridge CB1 4RN, U.K.

One of the more powerful methods by which to discover the functions of the proteolytic enzymes of living tissues depends on the use of specific inhibitors of these enzymes. We are successfully using inhibitory antisera in a study of the role of cathepsin D in cartilage breakdown (Dingle *et al.*, 1971), but effective inhibitors of low molecular weight might have important additional applications. Cathepsin D is a member of the 'acid proteinase' group of enzymes in which the hydrolysis of peptide bonds is thought to be catalysed by two or more carboxyl groups. Other enzymes in this group are another tissue proteinase, namely cathepsin E, the gastric enzymes pepsin (several forms), gastricsin and rennin, and numerous enzymes secreted by moulds and at least one bacterium. The acid proteinases most rapidly attack peptide bonds linking amino acids with hydrophobic side chains. Among the few known chemical inhibitors of these enzymes the best is probably diazoacetylnorleucine methyl ester in the presence of Cu^{2+} (Šodek & Hofmann, 1968; Lundblad & Stein, 1969; Keilová & Lapresle, 1970), but this often reacts slowly or incompletely, is unstable and probably is not highly specific. Certainly, it is not suitable for specific inhibition in biological systems, in the way that specific antisera to cathepsin D can be. The discovery of pepstatin in the culture medium of a strain of *Streptomyces* may therefore represent an important new development. Pepstatin has the structure isovaleryl-L-valyl-L-valyl-4-amino-3-hydroxy-6-methylheptanoyl-L-alanyl-4-amino-3-hydroxy-6-methylheptanoic acid (Morishima *et al.*, 1970), and its discoverers have reported that it inhibits pepsin and gastricsin (Umezawa *et al.*, 1970; Aoyagi *et al.*, 1971) and also acid proteinase activity attributed to cathepsin D in pig liver (Ikezawa *et al.*, 1971) and rabbit macrophages (Dannenburg *et al.*, 1972). The purpose of the experiments now reported was to establish whether this might be a suitable agent with which to investigate the biological functions of the tissue acid proteinases.

Materials

Pepstatin sodium salt, a product of the Banyu Pharmaceutical Co. Ltd., Tokyo, Japan, was kindly given by Dr. A. Dannenberg. Cathepsin D from liver of rabbit and man was prepared essentially as described by Barrett (1970); for each species an unresolved mixture of isoenzymes was used. A specific antiserum to rabbit cathepsin D was prepared as described by Dingle *et al.* (1971). Cathepsin B1 was isolated from human liver (A. J. Barrett, unpublished work). Trypsin (bovine, 2×crystallized; Sigma Chemical Co., St. Louis, Mo., U.S.A.), α-chymotrypsin (bovine, 3 × crystallized; Sigma), rennin (from calf stomach; Koch–Light Laboratories, Colnbrook, Bucks., U.K.), papain (2×crystallized; Sigma) and pepsin (2 × crystallized; Sigma) were purchased.

Results and discussion

Specificity of inhibition. The specificity of inhibition of proteolytic enzymes by pepstatin (5 μg/ml) was examined with trypsin (12.5 μg/ml, with α-*N*-benzoyl-DL-arginine 2-naphthylamide as substrate at pH7.5), chymotrypsin (50 μg/ml, with α-*N*-glutaryl-DL-phenylalanine 2-naphthylamide as substrate at pH7.5), papain (0.75 μg/ml, with α-*N*-benzoyl-DL-arginine 2-naphthylamide as substrate at pH6.0 in the presence of 2mM-cysteine and 1mM-EDTA) and cathepsin B1 (10 μg/ml, under conditions identical with those used for papain). Pepsin (6 μg/ml), cathepsin D (both rabbit and human) (8 μg/ml) and rennin (1 mg/ml) were assayed with haemoglobin as substrate (Barrett, 1970) at pH3.5, with 1 μg of pepstatin/ml.

Pepstatin had no effect on the activities of trypsin, chymotrypsin, papain or cathepsin B1, whereas pepsin, cathepsin D (from both species) and rennin were totally inhibited. The inhibition of cathepsin D was not diminished when the inhibitor had first been mixed with a large amount of heat-denatured (100°C, 5min) cathepsin D. Complete inhibition was also produced by pepstatin at 1 μg/ml (360ng/unit), whether or not haemoglobin (20mg/ml) had first been mixed with the enzyme, at pH3.5.

Since cathepsin E was not available to use in pure form, its sensitivity to inhibition was determined by use of a homogenate of rabbit bone marrow, a tissue known to contain cathepsin E (Lapresle, 1970). Portions (3.0ml) of the homogenate (1g of tissue in 20ml of 1% NaCl, pH6.8) were treated with sheep anti-(rabbit cathepsin D) serum (0–300 μl), stored
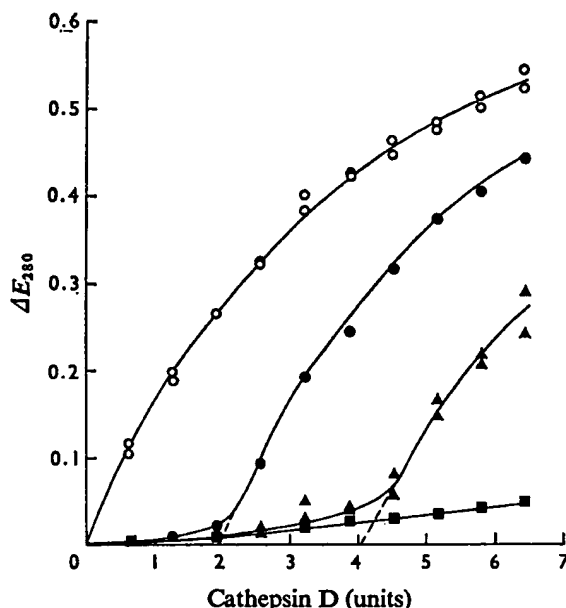
Fig. 1. *Effect of increasing cathepsin D concentration on the digestion of haemoglobin in the presence of pepstatin*

The digestion of haemoglobin was measured as the increase in E_{280} of trichloroacetic acid-soluble products in the absence of inhibitor (o) or in the presence of 50ng (●), 100ng (▲) or 200ng (■) of pepstatin. Each assay was made in a total volume of 1.0ml at pH3.5 (as described by Barrett, 1970), the enzyme and inhibitor being mixed at this pH. The results were obtained from two separate experiments.

overnight at 4°C and centrifuged, to remove cathepsin D. The proteolytic activity of the supernatant fluids (0.5ml) was determined at pH3.5, with haemoglobin as substrate, in the presence or absence of pepstatin (200ng/ml). It was found that 48% of the acid proteinase activity was not precipitable by the antiserum and this was attributed to the immunologically distinct cathepsin E. The entire acid proteinase activity of the extract was inhibited by pepstatin, however, and it was therefore concluded that cathepsin E is susceptible to inhibition by pepstatin.

The finding that pepstatin had no effect on the 'serine' or 'cysteine' proteinases, but was a potent inhibitor of the four acid proteinases tested, confirms and extends the results obtained by Aoyagi *et al.* (1971).

Kinetics of inhibition of cathepsin D. A method described by Ackerman & Potter (1949) (see also Khoo & Russell, 1970) was used to determine whether the inhibition of human cathepsin D by pepstatin is reversible, 'pseudo-irreversible' or

irreversible (Fig. 1). The results are characteristic of pseudo-irreversible or 'tight-binding' inhibition, in that the curves obtained in the presence of inhibitor tend to parallel the control curve obtained in its absence at high enzyme concentrations, but approach the origin, rather than intersecting the abscissa, at low concentrations of the enzyme. It is shown by extrapolation that, in the presence of excess of enzyme, 25ng of pepstatin completely inhibits 1 unit of human cathepsin D. On the basis that the specific activity of pure human cathepsin D is 600 units/mg and its molecular weight 45000 (Barrett, 1970), and that the molecular weight of sodium pepstatin is 708 (from Umezawa *et al.*, 1970), and assuming that the pepstatin preparation was pure, 25μg of pepstatin/unit of enzyme represents a molar ratio of 0.95:1.0. Equimolar binding of pepstatin to pepsin has been reported by Aoyagi *et al.* (1971). Our results do not permit an accurate calculation of the dissociation constant, but it appears to be less than 10^{-8} M. This finding is comparable with the value of less than 3×10^{-9} M for pepstatin with pepsin (Aoyagi *et al.*, 1971).

As would be expected from the extremely small dissociation constant, the inhibition of cathepsin D by pepstatin is very difficult to reverse experimentally; thus incomplete recovery of enzymic activity was achieved during dialysis for several days against a large volume of buffer.

Samples of human cathepsin D without inhibitor, or with 15, 30, 60 or 120ng of pepstatin/unit of enzyme, were subjected to analytical isoelectric focusing in polyacrylamide gel (essentially by the method of Barrett, 1970), and the gels were split longitudinally so that half could be stained for protein and the corresponding profile of enzymic activity could be determined by assays made with sections cut transversely from the other half. The samples containing pepstatin showed the same degree of inhibition as had been determined before isoelectric focusing, and there was no detectable difference in degree of inhibition between the several separated isoenzymes. No shift in the positions of the bands (attributable to the binding of pepstatin, a monobasic acid) was detected, even at the highest concentrations of inhibitor.

The antigenic properties of rabbit cathepsin D were shown quantitatively to be unaltered by the presence of a high concentration of pepstatin, in that the inhibitor did not affect the characteristics of binding of ^{125}I-labelled enzyme to an antibody attached covalently to Sepharose 4B (Dr. Zena Werb, personal communication).

It seems most probable that pepstatin has a very high affinity for the active site of acid proteinases. The predominantly hydrophobic nature of its side chains is compatible with the affinity of the acid proteinases for substrates in which the bond to be hydrolysed is flanked by non-polar residues. From the data of

Aoyagi *et al.* (1971) it seems that the hydroxyl group of the non-terminal residue of 4-amino-3-hydroxy-6-methylheptanoic acid is of great functional importance in the binding of pepstatin, perhaps through the formation of a hydrogen bond.

On the basis of the above results, pepstatin seems to be a valuable group specific inhibitor for the acid proteinases. It produces complete inhibition of cathepsins D and E at low concentration, binding in equimolar ratio to cathepsin D. In view of the high potency and low toxicity (Umezawa *et al.*, 1970) of pepstatin, it may well be of use in establishing the function of tissue acid proteinases in tissue systems.

Ackerman, W. W. & Potter, V. R. (1949) *Proc. Soc. Exp. Biol. Med.* 72, 1–9

Aoyagi, T., Kunimoto, S., Morishima, H., Takeuchi, T. & Umezawa, H. (1971) *J. Antibiot.* 24, 687–694

Barrett, A. J. (1970) *Biochem. J.* 117, 601–607

Dannenburg, A. M., McAdoo, M. H. & James, S. P. (1972). *Fed. Proc. Fed. Amer. Soc. Exp. Biol.* in the press.

Dingle, J. T., Barrett, A. J. & Weston, P. D. (1971) *Biochem. J.* 123, 1–13

Ikezawa, H., Aoyagi, T., Takeuchi, T. & Umezawa, H. (1971) *J. Antibiot.* 24, 488–490

Keilová, H. & Lapresle, C. (1970) *FEBS Lett.* 9, 348–350

Khoo, J. C. & Russell, P. J. (1970) *Biochim. Biophys. Acta* 220, 239–243

Lapresle, C. (1970) in *Tissue Proteinases* (Barrett, A. J. & Dingle, J. T., eds.), chapter 6, pp. 135–155, North-Holland Publishing Co., Amsterdam

Lundblad, R. L. & Stein, W. H. (1969) *J. Biol. Chem.* 244, 154–160

Morishima, H., Takita, T., Aoyagi, T., Takeuchi, T. & Umezawa, H. (1970) *J. Antibiot.* 23, 263–265

Šodek, J. & Hofmann, T. (1968) *J. Biol. Chem.* 243, 450–451

Umezawa, H., Aoyagi, T., Morishima, H., Matsuzaki, M., Hamada, M. & Takeuchi, T. (1970) *J. Antibiot.* 23, 259–262
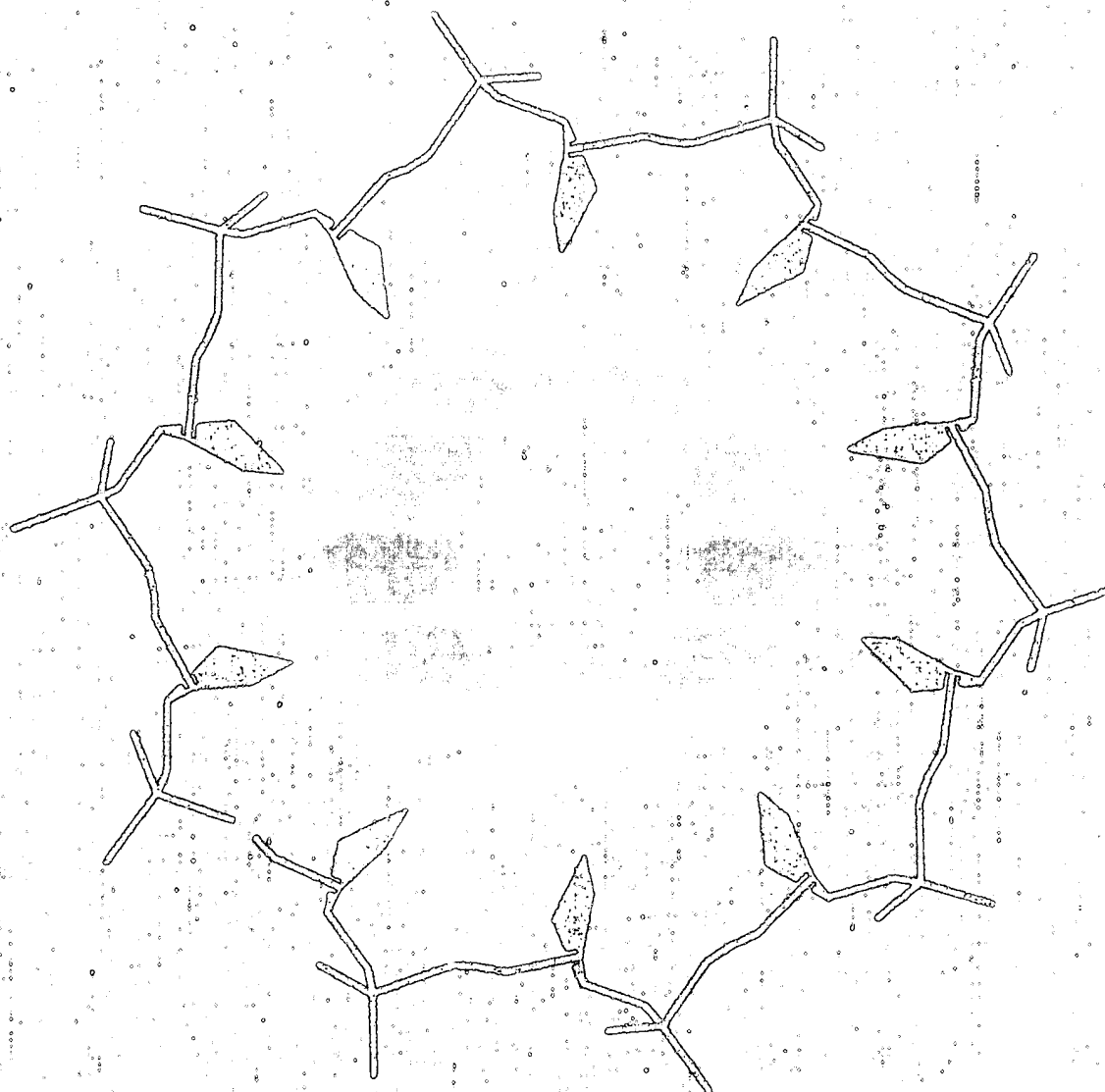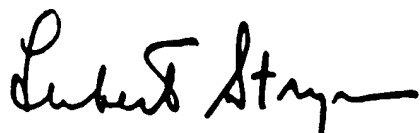
THIRD EDITION

BIOCHEMISTRY

THIRD EDITION

LUBERT STRYER

STANFORD UNIVERSITY

The need for substrate-induced structural changes in the active site of carboxypeptidase A can now be appreciated. The bound substrate is surrounded on all sides by catalytic groups of the enzyme. This arrangement promotes catalysis for the reasons cited above. It is evident that a substrate could not enter such an array of catalytic groups (nor could a product leave) unless the enzyme were flexible. *A flexible protein provides a much larger repertoire of potentially catalytic conformations than does a rigid one.*

SITE-SPECIFIC MUTAGENESIS IS A POWERFUL METHOD FOR ENGINEERING NEW ENZYMES AND ELUCIDATING MECHANISMS

In Agatha Christie's *Murder on the Orient Express*, Hercule Poirot's task as a detective was compounded by the presence of *so many* plausible suspects. Likewise, the catalytic mechanism of carboxypeptidase A has been difficult to unravel because of the presence of several potential catalytic groups at the active site. For example, the hydroxyl group of tyrosine 248 was postulated to be the proton donor to the NH group of the substrate in the cleavage step. This hypothesis was tested by William Rutter using site-specific mutagenesis. If tyrosine 248 is essential for catalysis, a mutant of this enzyme containing phenylalanine in its place should be inactive. The tyrosine 248 codon (TAT) was converted to that of phenylalanine (TTT) by oligonucleotide-directed mutagenesis (p. 136). The recombinant plasmid containing this gene was inserted into yeast and expressed. The striking finding was that the mutant enzyme had the same k_{cat} value as did the native enzyme but a K_M value that was sixfold higher. These results indicate that tyrosine 248 participates in the binding of substrate but is not essential for catalysis. This incisive experiment illustrates the power of site-specific mutagenesis in delineating the function of a particular residue in a protein.

THREE-DIMENSIONAL STRUCTURE OF CHYMOTRYPSIN, A SERINE PROTEASE

Chymotrypsin, like carboxypeptidase A, is a digestive enzyme. However, its specificity and catalytic mechanism are different, which makes it rewarding to examine chymotrypsin, too, in some detail. Chymotrypsin has further significance as a member of an important family of proteins, the *serine proteases*. Trypsin, another digestive enzyme, and thrombin, a clotting enzyme, are other members of this pervasive clan.

Chymotrypsin, a 25-kd enzyme, consists of three polypeptide chains connected by two interchain disulfide bonds. As will be discussed in the next chapter, chymotrypsin is synthesized as a single-chain inactive precursor called chymotrypsinogen. The three-dimensional structure of the enzyme was solved at 2-Å resolution (Figure 9-34) by David Blow. The molecule is a compact ellipsoid of dimensions $51 \times 40 \times 40$ Å. Chymotrypsin contains several antiparallel β-pleated sheet regions and little α helix. All charged groups are on the surface of the molecule except for three that play a critical role in catalysis.

The biological role of chymotrypsin is to catalyze the hydrolysis of proteins in the small intestine. Chymotrypsin does not cleave all peptide bonds at a significant rate. Rather, it is selective for peptide bonds on
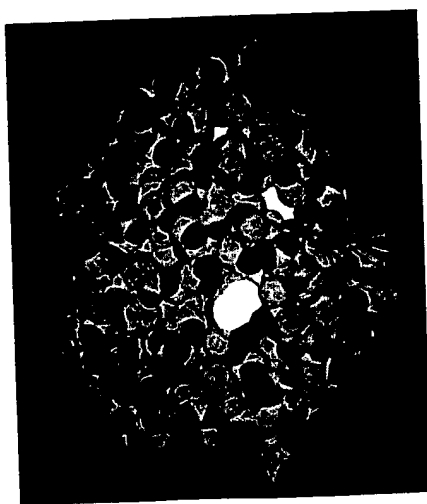
Space-filling model of chymotrypsin. This enzyme is stabilized by several disulfide bonds (yellow).

Figure 9-34
Three-dimensional structure of α-chymotrypsin. Only the α-carbon atoms are shown. Catalytically important residues are marked in color. [Courtesy of Dr. David Blow.]

the *carboxyl side* of the *aromatic side chains* tyrosine, tryptophan, and phenylalanine and of large *hydrophobic residues such as methionine*. Chymotrypsin also hydrolyzes *ester bonds*. Although unimportant physiologically, ester-bond hydrolysis is of interest because of its close relationship to peptide-bond hydrolysis (Figure 9-35). Indeed, much of our knowledge of the catalytic mechanism of chymotrypsin comes from studies of the hydrolysis of simple esters.
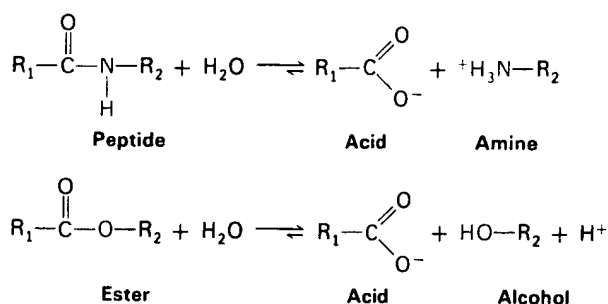
Figure 9-35
Chymotrypsin catalyzes the hydrolysis of peptide and ester bonds.

PART OF THE SUBSTRATE IS COVALENTLY BOUND TO CHYMOTRYPSIN DURING CATALYSIS

Chymotrypsin catalyzes the hydrolysis of peptide or ester bonds in two distinct stages. This was first revealed by studies of the kinetics of hydrolysis of *p*-nitrophenyl acetate. When large amounts of enzyme are used, there is an initial *rapid burst* of *p*-nitrophenol product, followed by its formation at a much *slower steady-state rate* (Figure 9-36).
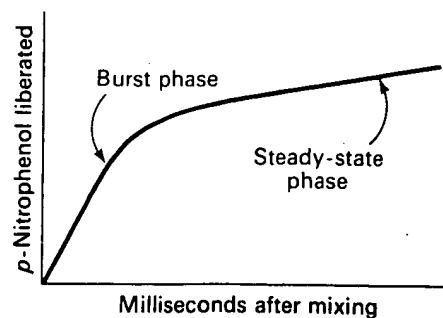
Figure 9-36
Two phases in the formation of *p*-nitrophenol are evident following the mixing of chymotrypsin and *p*-nitrophenyl acetate.
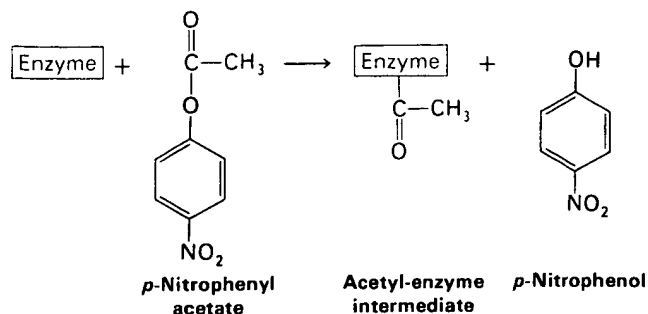
Figure 9-37
Acylation: formation of the acetyl-enzyme intermediate in
catalysis by chymotrypsin.

The first step is the combination of p-nitrophenyl acetate with chymotrypsin to form an enzyme-substrate (ES) complex (Figure 9-37). The ester bond of this substrate is cleaved. One of the products, p-nitrophenol, is then released from the enzyme, whereas the acetyl group of the substrate becomes covalently attached to the enzyme. Water then attacks the acetyl-enzyme complex to yield acetate ion and regenerate the enzyme (Figure 9-38). The initial rapid burst of p-nitrophenol pro-
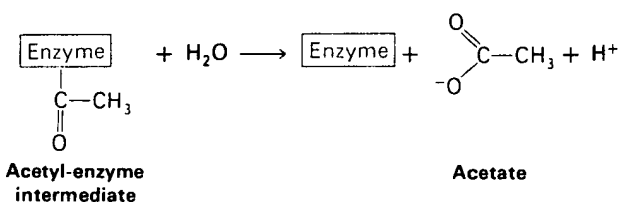
Figure 9-38
Deacylation: hydrolysis of the acetyl-enzyme intermediate.
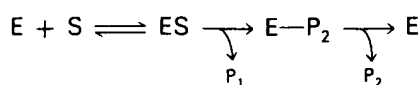
duction corresponds to the formation of the acetyl-enzyme complex. This step is called *acylation*. The slower steady-state production of p-nitrophenol corresponds to the hydrolysis of the acetyl-enzyme complex to regenerate the free enzyme. This second step, called *deacylation*, is much slower than the first, so that it determines the overall rate of hydrolysis of esters by chymotrypsin. In fact, the acetyl-enzyme complex is sufficiently stable to be isolated under appropriate conditions. The catalytic mechanism of chymotrypsin can thus be represented by the adjacent scheme, in which P_1 is the amine (or alcohol) component of the substrate, E—P_2 is the covalent intermediate, and P_2 is the acid component of the substrate.

$$E + S \rightleftharpoons ES \xrightarrow[P_1]{} E—P_2 \xrightarrow[P_2]{} E$$

A distinctive feature of this mechanism is the appearance of a covalent intermediate. In the particular reaction discussed above, an acetyl group is covalently bonded to the enzyme. In general, the group attached to chymotrypsin at the E—P_2 stage is an acyl group. Thus, E—P_2 is an *acyl-enzyme intermediate*.

THE ACYL GROUP IS ATTACHED TO AN UNUSUALLY REACTIVE SERINE RESIDUE ON THE ENZYME

The site of attachment of the acyl group was identified following the isolation of E—P_2, which is quite stable at pH 3. The acyl group is linked to the oxygen atom of a specific serine residue, namely, serine

195. This serine residue is unusually reactive. It can be specifically labeled with *organic fluorophosphates,* such as diisopropylphosphofluoridate (DIPF). DIPF reacts only with serine 195 to form an inactive *diisopropylphosphoryl-enzyme complex,* which is indefinitely stable. The remarkable reactivity of serine 195 is highlighted by the fact that the other 27 serine residues in chymotrypsin are untouched by DIPF.

Chymotrypsin is not the only enzyme to be inactivated by DIPF. Numerous other proteolytic enzymes, such as trypsin, elastase, thrombin, subtilisin, and acetylcholinesterase react specifically with DIPF and are inactivated. The reaction takes place at a unique serine residue, as in chymotrypsin. Hence, these enzymes are called the *serine proteases.*

$$\boxed{\text{Enzyme}} - CH_2 - O - \overset{\overset{\displaystyle O}{\|}}{C} - CH_3$$

Serine 195   Acetyl group

DEMONSTRATION OF THE CATALYTIC ROLE OF HISTIDINE 57 BY AFFINITY LABELING

The importance of a second residue in catalysis was shown by *affinity labeling.* The strategy was to react chymotrypsin with a molecule that (1) specifically binds to the active site because it resembles a substrate and then (2) forms a stable covalent bond with a group on the enzyme that is in close proximity. These criteria are met by tosyl-L-phenylalanine chloromethyl ketone (TPCK), whose structure is shown in Figure 9-39. The phenylalanine side chain of TPCK enables it to bind specifically to chymotrypsin. The reactive group in TPCK is the chloromethyl ketone function, which alkylates one of the ring nitrogens of histidine 57. TPCK is positioned to react with this residue because of its specific binding to the active site of the enzyme. The TPCK derivative of chymotrypsin is enzymatically inactive.

Three lines of evidence indicated that histidine 57 is part of the active site. First, the affinity-labeling reaction was highly stereospecific; the D-isomer of TPCK was totally ineffective. Second, the reaction was inhibited when a competitive inhibitor of chymotrypsin, β-phenylpropionate, was present. Third, the rate of inactivation by TPCK varied with pH in nearly the same way as did the rate of catalysis.

Ser 195

Diisopropylphosphoryl derivative of chymotrypsin

Specificity group

Reactive group

Tosyl-L-phenylalanine chloromethyl ketone (TPCK)

Chymotrypsin   + TPCK   Chymotrypsin
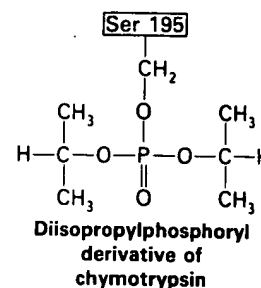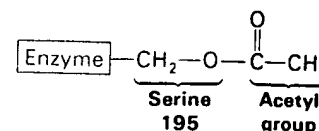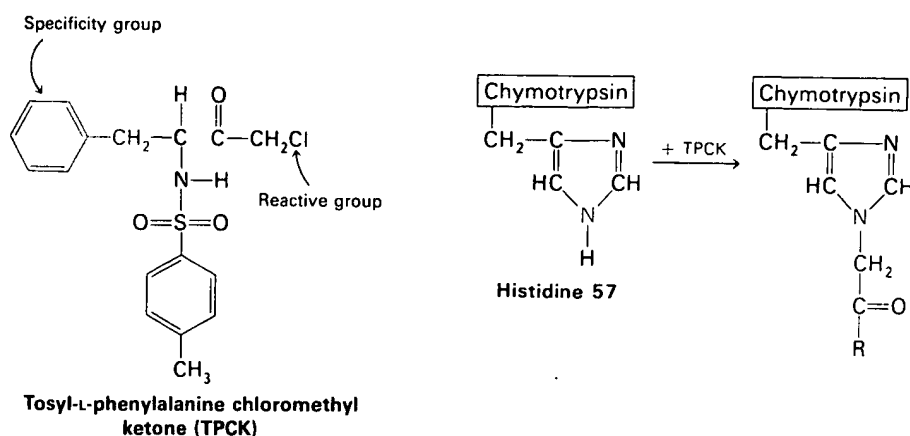
Histidine 57

Figure 9-39
Affinity labeling of a histidine residue in chymotrypsin by tosyl-L-phenylalanine chloromethyl ketone (TPCK), a reactive substrate analog.

SERINE, HISTIDINE, AND ASPARTATE FORM A CATALYTIC TRIAD IN CHYMOTRYPSIN

The catalytic activity of chymotrypsin depends on the unusual properties of serine 195. A —CH₂OH group is ordinarily quite unreactive under physiological conditions. What makes it so reactive in the active
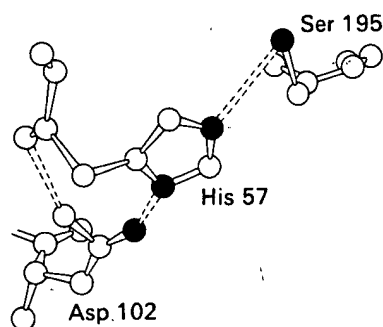
Figure 9-40
Conformation of the serine-
histidine-aspartate catalytic triad in
chymotrypsin. [After D. M. Blow
and T. A. Steitz. X-ray diffraction
studies of enzymes. *Ann. Rev. Bio-
chem.* 39(1970):86. Copyright © 1970
by Annual Reviews Inc. All rights
reserved.]

site of chymotrypsin? A convincing explanation has emerged from
x-ray studies of the three-dimensional structure of the enzyme. As was
foreseen by affinity-labeling studies, histidine 57 is adjacent to serine
195 (Figure 9-40). The carboxylate group of aspartate 102, buried in
the protein, also is next to histidine 57. These three residues form a
catalytic triad.

In the absence of substrate, histidine 57 is unprotonated (Figure
9-41). However, it is poised to accept the proton from the serine 195
—OH group when this oxygen atom carries out a nucleophilic attack on
the substrate. It was thought that aspartate 102 in turn became proton-
ated, and so this triad became known as a charge relay network. How-
ever, neutron diffraction studies have shown that the proton stays on
histidine 57 and that aspartate 102 remains negatively charged. In-
stead, the role of the —COO⁻ group of aspartate 102 is to stabilize the
positively charged form of histidine 57 in the transition state. In addi-
tion, aspartate 102 orients histidine 57 and insures that it is in the ap-
propriate tautomeric form to accept a proton from serine 195.

Figure 9-41
Role of the catalytic triad in chymotrypsin:
(A) enzyme alone; (B) on addition of a sub-
strate, a proton is transferred from serine
195 to histidine 57. The positively charged
imidazole ring is stabilized by electrostatic in-
teraction with negatively charged aspartate
102.
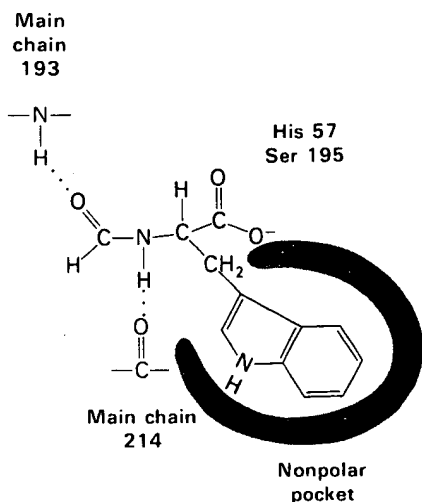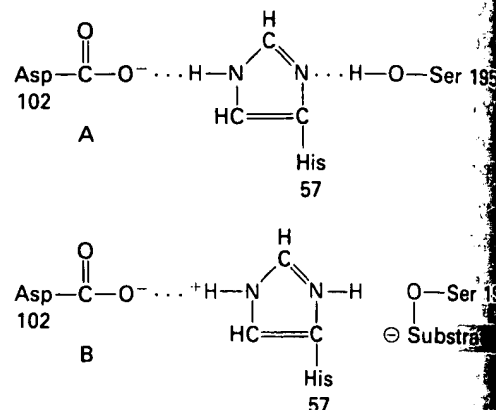
Figure 9-42
Schematic representation of the
binding of formyl-L-tryptophan, a
substrate analog, to chymotrypsin.

Crystallographic studies of complexes of chymotrypsin with substrate
analogs have also shown the location of the site of specific recognition
and the likely orientation of the susceptible peptide bond. Formyl-
tryptophan binds to chymotrypsin with its indole side chain fitting
neatly into a pocket near serine 195 (Figure 9-42). This deep cleft ac-
counts for the specificity of chymotrypsin for aromatic and other bulky
hydrophobic side chains. Crystallographic analyses of complexes of
chymotrypsin and polypeptide substrate analogs show that the main
chain of the enzyme and that of the substrate are hydrogen bonded to
each other as in an antiparallel β pleated sheet.

A TRANSIENT TETRAHEDRAL INTERMEDIATE
IS FORMED DURING CATALYSIS BY CHYMOTRYPSIN

A plausible catalytic mechanism for chymotrypsin has been deduced
from extensive x-ray crystallographic and chemical data. In this mecha-
nism, *histidine 57 and serine 195 participate directly in the cleavage of the
susceptible peptide bond of the substrate.* Hydrolysis of the peptide bond
starts with an attack by the oxygen atom of the hydroxyl group of serine

on the carbonyl carbon atom of the susceptible peptide bond. The carbon–oxygen bond of this carbonyl group becomes a single bond, and the oxygen atom acquires a net negative charge. The four atoms now bonded to the carbonyl carbon are arranged as in a tetrahedron. The formation of this *transient tetrahedral intermediate* from a planar amide group is made possible by hydrogen bonds between the negatively charged carbonyl oxygen atom (called an *oxyanion*) and two main-chain NH groups (Figure 9-43).

The other essential event in the formation of this tetrahedral transition state is the transfer of a proton from serine 195 to histidine 57 (Figure 9-44). This proton transfer is markedly facilitated by the presence of the catalytic triad. Aspartate 102 precisely orients the imidazole ring of histidine 57 and partly neutralizes the charge that develops on it during the transition state. The proton held by the protonated form of histidine 57 is then donated to the nitrogen atom of the susceptible peptide bond, which thus is cleaved. At this stage, the amine component is hydrogen bonded to histidine 57, whereas the acid component of the substrate is esterified to serine 195. The amine component diffuses away, completing the *acylation stage* of the hydrolytic reaction.

The next stage, *deacylation* (Figure 9-45), begins when a water molecule takes the place occupied earlier by the amine component of the substrate. In essence, *deacylation is the reverse of acylation, with H_2O substi-*
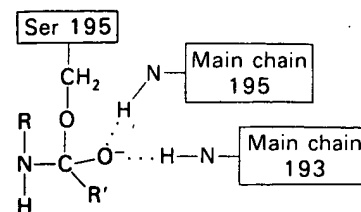
Figure 9-43
The tetrahedral transition state in the acylation reaction of chymotrypsin. The hydrogen bonds formed by two NH groups from the main chain of the enzyme are critical in stabilizing this species. This site is called the *oxyanion hole*.
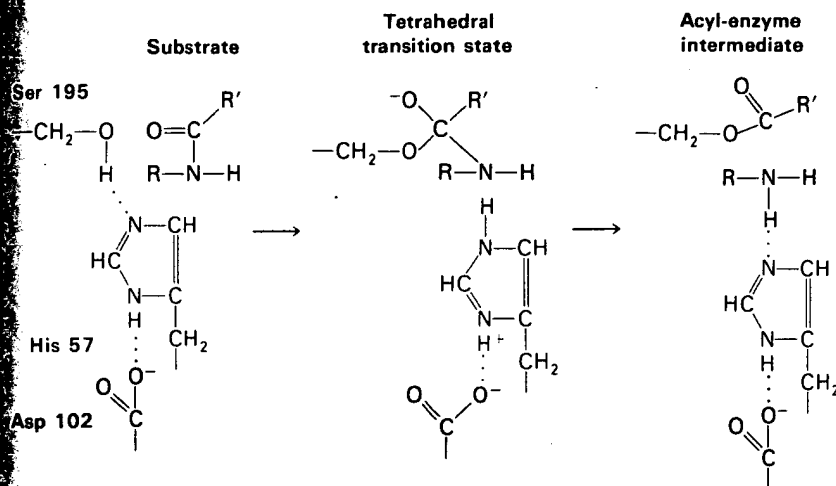
Figure 9-44
First stage in the hydrolysis of a peptide by chymotrypsin: *acylation*. A tetrahedral transition state is formed, in which the peptide bond is cleaved. The amine component then rapidly diffuses away, leaving an acyl-enzyme intermediate.
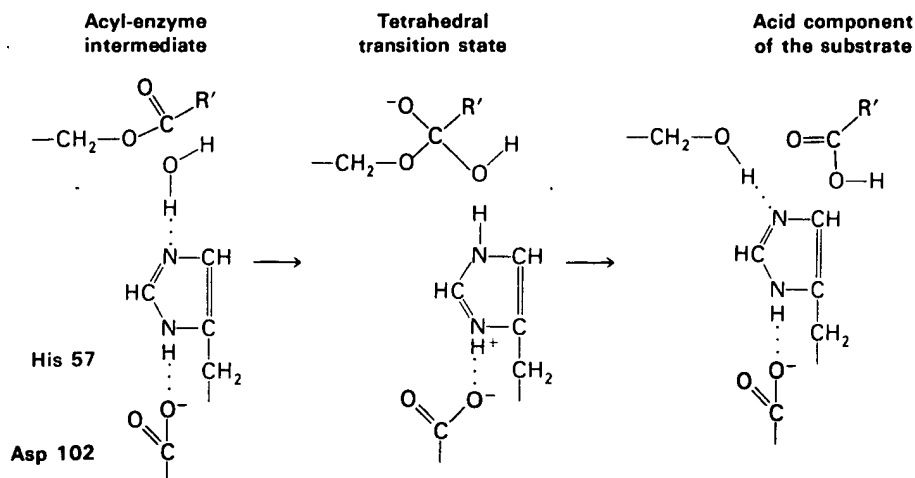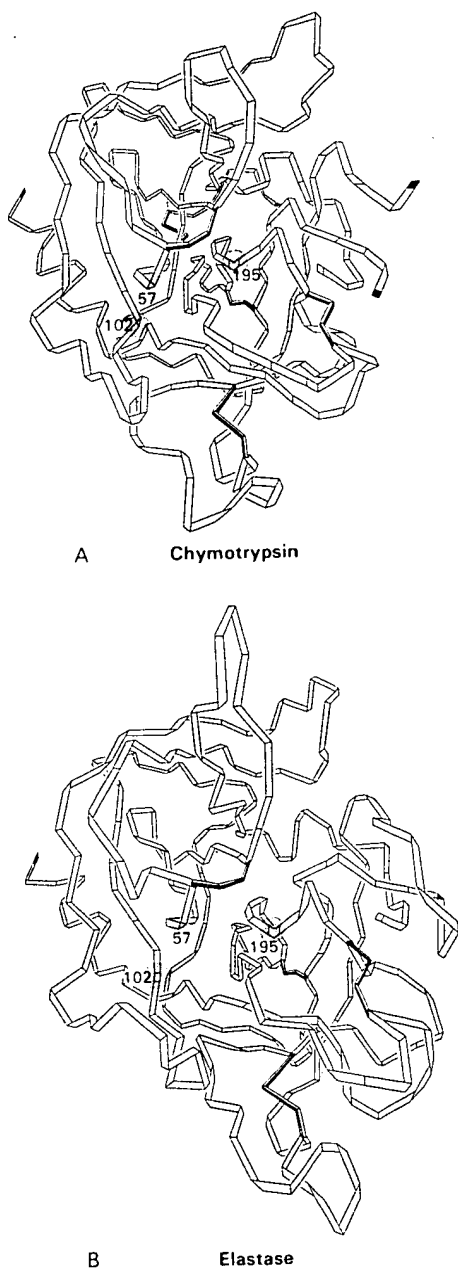
Figure 9-45
Second stage in the hydrolysis of a peptide by chymotrypsin: *deacylation*. The acyl-enzyme intermediate is hydrolyzed by water. Note that deacylation is essentially the reverse of acylation, with water in the role of the amine component of the original substrate.

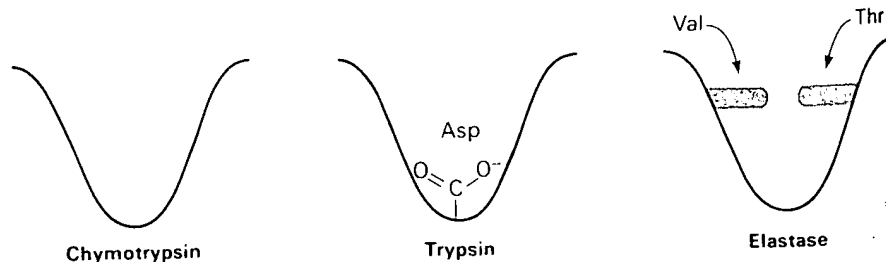A   Chymotrypsin

B   Elastase

Figure 9-46
Comparison of the conformation of
the main chains of (A) chymotrypsin
and (B) elastase. The locations of the
catalytic triad (residues 102, 57, and
195) are shown in color. [After B. S.
Hartley and D. M. Shotton. In *The
Enzymes*, P. D. Boyer, ed., 3rd ed.,
vol. 3 (Academic Press, 1971), p.
362.]

tuting for the amine component. First, the charge relay network draws a
proton away from water. The resulting OH⁻ ion immediately attacks
the carbonyl carbon atom of the acyl group that is attached to serine
195. As in acylation, a transient tetrahedral intermediate is formed.
Histidine 57 then donates a proton to the oxygen atom of serine 195,
which then releases the acid component of the substrate. This acid com-
ponent diffuses away and the enzyme is ready for another round of
catalysis.

TRYPSIN AND ELASTASE: VARIATIONS ON THE CHYMOTRYPSIN THEME

Trypsin and elastase, two other digestive enzymes, are like chymotryp-
sin in several respects: (1) About 40% of the amino acid sequences of
these three enzymes is identical. The degree of identity is even higher
(~60%) for residues located in the interior of the enzymes. (2) X-ray
studies have shown that their three-dimensional structures are very
similar (Figure 9-46). (3) A serine-histidine-aspartate catalytic triad is
present in all three. (4) The serine residue in this triad is modified by
fluorophosphates (such as DIPF), causing a loss of catalytic activity. The
amino acid sequence around this serine is the same in all three enzymes,
Gly-Asp-Ser-Gly-Gly-Pro. (5) All three enzymes have nearly identical
catalytic mechanisms. The catalytic triad and oxyanion hole in each
promotes the formation of a tetrahedral transition state. A covalent
acyl-enzyme intermediate is formed by all three during catalysis. (6) As
will be discussed in the next chapter, these enzymes are secreted by the
pancreas as inactive precursors that become activated by cleavage of a
single peptide bond.

Although similar in structure and mechanism, these enzymes differ
markedly in substrate specificity. Chymotrypsin requires an aromatic or
bulky nonpolar side chain on the amino side of the scissile bond. Tryp-
sin requires a lysine or arginine residue. Elastase cannot cleave either of
these kinds of substrates. Its specificity is directed toward the smaller
uncharged side chains. X-ray studies have shown that *these different spec-
ificities are due to quite small structural differences in the binding site* (Figure
9-47). In chymotrypsin, a nonpolar pocket serves as a niche for the
aromatic or bulky nonpolar side chain. In trypsin, one residue in this
pocket is different from chymotrypsin: a serine is replaced by an aspar-
tate. This aspartate in the nonpolar pocket of trypsin can form a strong
electrostatic bond with a positively charged lysine or arginine side chain.

Figure 9-47
A highly simplified representation of part of the substrate-binding site in
chymotrypsin, trypsin, and elastase.